

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
28 February 2002 (28.02.2002)

PCT

(10) International Publication Number  
**WO 02/17120 A2**

(51) International Patent Classification<sup>7</sup>: **G06F 17/00**

(21) International Application Number: PCT/US01/25753

(22) International Filing Date: 17 August 2001 (17.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/641,589 18 August 2000 (18.08.2000) US  
09/836,484 17 April 2001 (17.04.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US 09/836,484 (CON)

Filed on 17 April 2001 (17.04.2001)

US 09/641,589 (CON)

Filed on 18 August 2000 (18.08.2000)

(71) Applicant (for all designated States except US): **THINK-ING INVESTMENTS, INC.** [US/US]; Suite 555, 800 South Street, Waltham, MA 02453 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **COOPER, Philip**, A. [US/US]; 61 Buckskin Drive, Weston, MA 02493 (US). **COOPER, Lisette** [US/US]; 61 Buckskin Drive, Weston, MA 02493 (US). **MYERS, Stuart** [US/US]; 60 Meriam Road, Lexington, MA 02173 (US). **FORNEY, G., David, Jr.** [US/US]; 1010 Memorial Drive #3G, Cambridge, MA 02138 (US). **SCOTT, Leonard, L., Jr.** [US/US];

325 Tearose Lane, Charlottesville, MA 22903 (US). **SHECTMAN, Benjamin** [US/US]; 72 Abbotsford Road, Noth Plainfield, NJ 07062 (US). **LECLAIR, Raymond** [US/US]; -2-Sandas-Trail,-Acton,-MA-01722-(US)-**LI, Yongxiang** [CN/CN]; 11-B Hamre Lane, Brandford, CT 06405 (CN).

(74) Agent: **FEIGENBAUM, David, L.**; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: GENERATING AND PROVIDING INFORMATION ABOUT EXPECTED FUTURE PRICES OF ASSETS AND VISUALIZATION OF ASSET INFORMATION

(57) Abstract:

WO 02/17120 A2

**GENERATING AND PROVIDING INFORMATION ABOUT  
EXPECTED FUTURE PRICES OF ASSETS AND  
VISUALIZATION OF ASSET INFORMATION**

**BACKGROUND**

---

5     This invention relates to generating and providing information about  
expected future prices of assets, and to visualization of asset information.

Among the kinds of information available at web sites on the Internet are  
current and historical prices and volumes of stock transactions, prices of  
put or call options at specific strike prices and expiration dates for various  
10     stocks, and theoretical prices of put and call options that are derived using  
formulas such as the Black-Scholes formula. Some web sites give  
predictions by individual experts of the future prices or price ranges of  
specific stocks.

A call option gives the holder a right to buy an underlying marketable  
15     asset by an expiration date for a specified strike price. A put option gives  
an analogous right to sell an asset. Options are called derivative securities  
because they derive their values from the prices of the underlying assets.  
Examples of underlying assets are corporate stock, commodity stock, and  
currency. The price of an option is sometimes called the premium.

20     People who buy and sell options are naturally interested in what  
appropriate prices might be for the options. One well-known formula for  
determining the prices for call and put options under idealized conditions  
is called the Black-Scholes formula. Black-Scholes provides an estimate  
of call or put prices for options having a defined expiration date, given a  
25     current price of the underlying asset, an interest rate, and the volatility rate  
(sometimes simply called volatility) of the asset. Black-Scholes assumes

constant interest rates and volatility, no arbitrage, and trading that is continuous over a specified price range.

Information about investment assets such as corporate securities is often presented as tables of values or ratios of values for successive time periods.

Sometimes graphs or visualization devices are used to provide a more intuitive view of the information.

One on-line service, Morningstar.com, uses a scatter plot in its Morningstar Investment Radar, URL

(<http://screen.morningstar.com/InvestmentRadar/InvestmentRadar.html>). Each point in the plot represents risk versus capitalization of an asset in a portfolio.

Another on-line facility, FalconEye, URL

(<http://www.falconeye.com/falconeye/tracker/index.html>), displays a periscope-like view of a simulated cloud formation that represents

a multi-dimensional density map of all 6000+ Nasdaq stocks, sorted in real-time by FalconEye Viz-Alerts™ (customizable indicators) that were created for the vertical and horizontal axes. Each stock is like a pixel on the screen and each color represents the density of stocks depicted in that section of the Tracker Live Map. ... The distribution of density allows you to instantly see the real-time technical pressures on the market and gives you the knowledge to trade more efficiently and productively.

ValuEngine, URL (<http://valuengine.com/servlet/ValuationSummary#>), displays graphs of stock prices that include historical prices to a current date followed by forecast price trends for future periods, including forecast ranges above and below the forecast price trends.

**SUMMARY**

In general, in one aspect, the invention features a method in which data is received that represents current prices of options on a given asset. An estimate is derived from the data of a corresponding implied probability

5 distribution of the price of the asset at a future time. Information about the probability distribution is made available within a time frame that is useful to investors, for example, promptly after the current option price information becomes available.

Implementations of the invention may include one or more of the  
10 following features. The data may represent a finite number of prices of options at spaced-apart strike prices of the asset. A set of first differences may be calculated of the finite number of prices to form an estimate of the cumulative probability distribution of the price of the asset at a future  
15 time. A set of second differences may be calculated of the finite number of strike prices from the set of first differences to form the estimate of the probability distribution function of the price of the asset at a future time.

In general, in another aspect, the invention features a method in which a real time data feed is provided that contains information based on the probability distribution.

20 In general, in another aspect, the invention features a method that includes providing a graphical user interface for viewing pages containing financial information related to an asset; and when a user indicates an asset of interest, displaying probability information related to the price of the asset at a future time.

25 In general, in another aspect, the invention features a method that includes receiving data representing current prices of options on a given asset, the options being associated with spaced-apart strike prices of the asset at a

Attorney Docket 11910-002001

future time. The data includes shifted current prices of options resulting from a shifted underlying price of the asset, the amount by which the asset price has shifted being different from the amount by which the strike prices are spaced apart. An estimate is derived from a quantized implied probability distribution of the price of the asset at a future time, the elements of the quantized probability distribution being more finely spaced than for a probability distribution derived without the shifted current price data.

10 In general, in another aspect, the invention includes deriving from said data an estimate of an implied probability distribution of the price of the asset at a future time, the mathematical derivation including a smoothing operation.

15 Implementations of the invention may include one or more of the following features. The smoothing operation may be performed in a volatility domain.

In general, in another aspect, the invention includes deriving a volatility for each of the future dates in accordance with a predetermined option pricing formula that links option prices with strike prices of the asset; and generating a smoothed and extrapolated volatility function.

20 Implementations of the invention may include one or more of the following features. The volatility function may be extrapolated to a wider range of dates than the future dates and to other strike prices. The smoothed volatility function may be applicable to conditions in which the data is reliable under a predetermined measure of reliability. The implied volatility function formula may have a quadratic form with two variables  
25 representing a strike price and an expiration date. The coefficients of the implied volatility function formula may be determined by applying

Attorney Docket 11910-002001

regression analysis to approximately fit the implied volatility function formula to each of the implied volatilities.

In general, in another aspect, the invention features a method that includes receiving data representing current prices of options on assets belonging to

5 a portfolio, deriving from the data an estimate of an implied multivariate distribution of the price of a quantity at a future time that depends on the assets belonging to the portfolio, and making information about the probability distribution available within a time frame that is useful to investors.

10 In general, in another aspect, the invention features a method that includes receiving data representing values of a set of factors that influence a composite value, deriving from the data an estimate of an implied multivariate distribution of the price of a quantity at a future time that depends on assets belonging to a portfolio, and making information about  
15 the probability distribution available within a time frame that is useful to investors.

Implementations of the invention may include one or more of the following features. The mathematical derivation may include generating a multivariate probability distribution function based on a correlation among  
20 the factors.

In general, in another aspect, the invention features a graphical user interface that includes a user interface element adapted to enable a user to indicate a future time, a user interface element adapted to show a current price of an asset, and a user interface element adapted to show the  
25 probability distribution of the price of the asset at the future time.

In general, in one aspect, the invention features, a method that includes continually generating current data that contains probability distributions of prices of assets at future times, continually feeding the current data to a recipient electronically, and the recipient using the fed data for services  
5 provided to users.

In general, in another aspect, the invention features a method that includes receiving data representing current prices of options on assets belonging to a portfolio, receiving data representing current prices of market  
10 transactions associated with a second portfolio of assets, and providing information electronically on the probability that the second portfolio of assets will reach a first value given the condition that the first portfolio of assets reaches a specified price at a future time.

In general, in another aspect, the invention features a method that includes receiving data representative of actual market transactions associated with  
15 a first portfolio of assets; receiving data representative of actual market transactions associated with a second portfolio of assets; and providing information on the expectation value of the price of first portfolio of assets given the condition that the second portfolio of assets reach a first specified price at a specified future time through a network.

20 In general, in another aspect, the invention features a method that includes evaluating an event defined by a first multivariate expression that represents a combination of macroeconomic variables at a time T, and estimating (e.g., using Monte Carlo techniques) the probability that a second multivariate expression that represents a combination of values of  
25 assets of a portfolio will have a value greater than a constant B at time T if the value of the first multivariate expression is greater than a constant A. The market variables represented by the first multivariate expression can

include macroeconomic factors (such as interest rates), market preferences regarding the style of company fundamentals (large/small companies, rapid/steady growth, etc.), or market preferences for industry sectors.

In general, in another aspect, the invention features a method that includes

- 5 defining a regression expression that relates the value of one variable representing a combination of macroeconomic variables at time T to a second variable at time T that represents a combination of assets of a portfolio, and estimating the probability that the second variable will have a value greater than a constant B at time T if the value of the first variable  
10 is greater than a constant A at time T, based on the ratio of the probability of x being greater than A under the regression expression and the probability of x being greater than A.

- In general, in another aspect, the invention features a method that includes defining a current value of an option as a quadratic expression that  
15 depends on the difference between the current price of the option and the current price of the underlying security, and using Monte Carlo techniques to estimate a probability distribution of the value at a future time T of a portfolio that includes the option.

- The invention takes advantage of the realization that option prices for a  
20 given underlying asset are indicative of the market's prediction of the of the risk-neutral price of the underlying asset in the future (e.g., at the expiration of the option). Option price data may be used to derive the market's prediction in the form of an implied probability distribution of future risk-neutral prices. Additional explanation of the significance of the  
25 phrase risk-neutral is contained in the Appendix.

The implied probability distribution and other information related to it may be made easily available to people for whom the information may be



useful, such as those considering an investment in the underlying asset, or a brokerage firm advising such an investor.

In general, in another aspect, the invention features a method that includes (a) displaying to a user a circular visualization element having sectors

5 arranged around a center of the element, the sectors respectively corresponding to different groups of assets, and (b) in each of the sectors, displaying an array of visual elements representative of respective assets belonging to the group to which the sector corresponds, the visual  
10 elements being arrayed with respect to distance from the center in accordance with magnitudes of performance of the assets during a recent period.

Implementations of the invention may include one or more of the following features. The visual elements comprise displayed dots, one for each of the assets. The visual elements exhibit visible characteristics that  
15 correspond to categories of the assets within the group. The categories of the assets within the group correspond to different capitalizations. The dots are arranged along a radius of the sector to which they belong. Dots that would otherwise lie on the radius at a given distance from the center are displayed at different angular positions near to the radius. Each sector  
20 has an angular extent that represents the fraction of asset items in the sector relative to the total number of asset items in the universe being plotted. The circular visualization element is subdivided into rings having respectively different distances from the center. The rings are displayed in different colors. The magnitudes of performance of the assets are  
25 measured in percentage price change. The recent period comprises a trading day on an asset market. The assets comprise securities issued by corporations.

In general, in another aspect, the invention features a method that includes displaying to a user a visualization element that indicates the odds of a performance measure of an asset being within specified ranges of identified values of the performance measure at a succession of times in the future.

Implementations of the invention may include one or more of the following features. The performance measure comprises a price of the asset or a return percentage or a tax-adjusted return percentage. The visualization element includes stripes superimposed on a graph of the performance measure over time, each of the stripes representing one of the specified ranges. Each of the stripes begins at a current time and becomes broader as it extends to future times. A graphical device shows actual historical values of the performance measure, e.g. in the form of a line graph one end of which joins the visualization element at a point that represents a current date. The visualization element includes two portions, one of the portions representing the odds prior to a specified date based on one assumption, the other of the portions representing the odds after the specified date based on another assumption. The specified date is a date on which tax effects change from the one assumption to the other assumption.

In general, in another aspect, the invention features a method that includes displaying to a user a visualization element having graphical indicators of the relative performance of a selected asset compared with the performance of groups of assets in each of a succession of time periods, each of the groups comprising assets representing a common style. The relative performance is determined using an asset class factor model.

Among the advantages of the invention are one or more of the following: Investors and prospective investors in an underlying asset, such as a

publicly-traded stock, are given access to a key additional piece of current information, namely calculated data representing the market's view of the future price of the stock. Brokerage firms, investment advisors, and other companies involved in the securities markets are able to provide the  
5 information or related services to their clients and customers. A user is enabled to quickly visualize and grasp the significance of data that would otherwise be more difficult to understand.

Other features and advantages will become apparent from the following description and from the claims.

## 10 DESCRIPTION

Details of implementations of the invention are set forth in the figures and the related description below.

Figures 1, 2, and 3 are graphs.

Figure 4 is a block diagram.

15 Figures 5, 6, and 7 are web pages.

Figures 8 and 9 illustrate user interfaces.

Figure 10 shows data structures.

Figures 11 through 15 show visualization techniques.

20 In general, the price of a call or put option is determined by buyers and sellers in the option market and carries information about the market's prediction of the expected price of the underlying asset at the expiration date. (The information does not include the premium that investors require for bearing risk, which must be estimated separately. The average long-

term value of the risk premium is about 6% per year for all stocks and may be adjusted for an individual stock's historical responsiveness to broader market movements.)

The information carried in the prices of options having various strike

- 5 prices and expirations is used to derive probability distributions of the asset's price at future times and to display corresponding information to investors, for example, on the World Wide Web.

*Basic method*

- 10 We first define some relevant terms. We define  $x$  as the strike price,  $c(x)$  as the theoretical call price function (the price of the call as a function of strike price),  $p(x)$  as the theoretical put price function,  $F(x)$  as the cumulative distribution function (cdf) of the price of the underlying asset at expiration; and  $f(x)$  as the probability density function (pdf) of the asset price at expiration. By definition,  $f(x) = F'(x)$  (i.e., the probability density  
15 function is the derivative of the cumulative distribution function).

The relationship between  $c(x)$ ,  $p(x)$ ,  $f(x)$ , and  $F(x)$  can be succinctly stated as:

$$F(x) = c'(x) + 1 = p'(x); \quad (1a)$$

$$f(x) = c''(x) = p''(x). \quad (1b)$$

- 20 In words, the pdf is the second derivative of either the call price function or of the put price function. A simple proof of these relationships is given in the Appendix. The Appendix also contains other detailed information relating to the features of the invention.

- 25 This so-called "second-derivative method" for computing implied probability distributions from option price data is known in the academic literature, but apparently not very well known. For example, the standard

textbook "Options, Futures, and Other Derivatives," by John C. Hull (Fourth Edition, 1999; Prentice-Hall) mentions implied probabilities, but not the second-derivative method. Perhaps the best reference that we have been able to find is J. C. Jackwerth and M. Rubinstein, "Recovering  
 5 probability distributions from option prices," J. Finance, vol. 51, pp. 1611-1631 (1996), which has only six prior references. This paper cites D. T. Breeden and R. H. Litzenberger, "Prices of state-contingent claims implicit in option prices," J. Business, vol. 51, pp. 631-650 (1978) as the originator of a second-derivative method, although the latter paper  
 10 nowhere mentions probabilities.

*Approximating  $f(x)$  from finite bid and ask option prices*

Equations (1a) and (1b) are obtained by assuming that the variable  $x$  is continuous and ranges from 0 to infinity. In practice, options are usually traded within certain price ranges and only for certain price intervals (e.g.,  
 15 ranging from \$110 to \$180 at \$5 intervals). Thus, the call and/or put option prices are known only for a finite subset of strike prices. Under such circumstances, estimates of Equations (1a) and (1b) can be computed by taking differences instead of derivatives as follows.

We assume that the option prices  $c(x)$  and  $p(x)$  are quoted for a finite  
 20 subset of equally-spaced strike prices  $x = n \Delta$ , where  $n$  is an integer, and  $\Delta$  is the spacing between quoted prices. Define  $c_n = c(n \Delta)$ ,  $p_n = p(n \Delta)$ . Then the first derivatives  $c'(x)$  and  $p'(x)$  at  $x = (n + \frac{1}{2})\Delta$  may be estimated by the first differences:

$$\hat{c}'_{n+1/2} = \frac{c_{n+1} - c_n}{\Delta}; \quad (7a)$$

$$25 \quad \hat{p}'_{n+1/2} = \frac{p_{n+1} - p_n}{\Delta}. \quad (7b)$$

The corresponding estimates of the cumulative distribution function:

$$\hat{F}_{n+1/2} = F((n + \frac{1}{2})\Delta) \text{ are}$$

$$\hat{F}_{n+1/2} = 1 + \hat{c}'_{n+1/2} \quad (8a)$$

$$\hat{F}_{n+1/2} = \hat{p}'_{n+1/2} \quad (8b)$$

- 5 The second derivatives  $c''(x)$  and  $p''(x)$  at  $x = n \Delta$  may likewise be estimated by the second differences, i.e., differences of the estimates of the first derivatives:

$$\hat{c}''_n = \frac{c_{n+1/2} - c_{n-1/2}}{\Delta} = \frac{c_{n+1} - 2c_n + c_{n-1}}{\Delta^2}, \quad (9a)$$

$$\hat{p}''_n = \frac{p_{n+1} - 2p_n + p_{n-1}}{\Delta^2}. \quad (9b)$$

- 10 Either of these estimates of the second derivatives may be used as an estimate of the probability density values at  $x = n \Delta$ , i.e.,  $f(n\Delta)$ :

$$\hat{f}_n = \hat{p}''_n \text{ or } \hat{f}_n = \hat{c}''_n \quad (10)$$

- Moreover, the market prices of call and put options are usually given in terms of a bid-ask spread, and thus either the bid price or the ask price (or some intermediate value) may be used as the call or put option price. By using the bid and ask prices for both the call option and the put option, four estimates of  $F(x)$  and  $f(x)$  may be obtained. These estimates may be combined according to their reliability in any desired way. For example, one might use the estimate derived from the put bid price curve for values of  $x$  less than the current price  $s$  of the underlying asset, and the estimate derived from the call bid price curve for values of  $x$  greater than  $s$ .

Examples of  $c_n$ ,  $p_n$ ,  $\hat{F}_{n+1/2}$ , and  $\hat{f}_n$  are shown in figures 1, 2, and 3 using the data of TABLE 1 (see below).

*Tabular data*

TABLE 1 below shows sample bid prices of call and put options for strike  
5 prices of an asset ranging from \$110 to \$180 at \$5 intervals and the  
cumulative distribution values  $\hat{F}_{n+1/2}$  and probability density values  $\hat{f}_n$   
computed according to Equations (7)-(10) above.

In the table, the values for  $\hat{F}_{n+1/2}$  correspond to strike prices that are mid-  
way between the two strike prices used to compute  $\hat{F}_{n+1/2}$ . Thus, the  
10 cumulative distribution value shown to the right of the strike price \$110  
actually corresponds to the strike price \$112.5, and the value to the right  
of the strike price \$115 actually corresponds to strike price \$117.5, and so  
forth.

Strike price	Call price	$\hat{F}_{n+1/2}$ from call price	$\hat{f}_n$ from call price	Put price	$\hat{F}_{n+1/2}$ from put price	$\hat{f}_n$ from put price
110	42 7/8	0	0	1/	0	0
115	37 7/8	0.025	0.025	1/	0.0125	0.0125
120	33	0.025	0	3/1	0.0125	0
125	28 1/8	0.05	0.025	1/	0.0375	0.025
130	23 3/8	0.125	0.075	7/1	0.0875	0.05
135	19	0.175	0.05	7/	0.15	0.0625
140	14 7/8	0.225	0.05	1 5/	0.2375	0.0875
145	11	0.35	0.125	2 13/1	0.3875	0.15
150	7 3/4	0.525	0.175	4 3/	0.5	0.1125
155	5 3/8	0.6	0.075	7 1/	0.6	0.1
160	3 3/8	0.7375	0.1375	10 1/	0.725	0.125
165	2 1/16	0.825	0.0875	13 7/	0.825	0.1
170	1 3/16	0.8875	0.0625	1	0.85	0.025
175	5/8	0.925	0.0375	22 1/	0.925	0.075
180	1/4			26 7/		

Dynamic estimates for  $F(x)$  and  $f(x)$ .

In Equations (7)-(10), the call and put option prices were assumed to be static in the calculation of the cumulative distribution function  $F(x)$  and probability density function  $f(x)$  for a finite subset of strike prices  $x = n \Delta$ .

In the real world, the price  $s$  of the underlying asset changes with time, and there will be a corresponding change in option prices. As a first order approximation, if the price  $s$  increases by a small amount  $\delta$ , then the option price curves will effectively shift to the right by the amount  $\delta$ .

(Here,  $\delta$  may be either positive or negative. For a more precise discussion of the shift, see the Appendix.) As a result, the price  $c(x)$  or  $p(x)$  now quoted at strike price  $x$  may be used as an estimate for the option price on



the previous price curve at strike price  $x' = x - \delta$ . As a result, the prices on the previous curve at a new discrete subset of strike prices  $x = n\Delta - \delta$  become effectively visible. Given enough movements of the underlying price, therefore, we can effectively compute estimates of  $c(x)$ ,  $p(x)$ ,  $F(x)$  and/or  $f(x)$  for a subset of strike prices  $x$  that is much more closely spaced than the subset available at any one time.

*Extrapolating and smoothing probability distributions.*

In a typical options market, the option prices are available only for certain expiration dates. In addition, the option prices are more reliable for options that are actively traded, which are typically nearer-term options at strike prices near the underlying price. It is therefore desirable to extrapolate and interpolate probability distributions to times other than actual expiration dates and to wider ranges of strike prices.

Any standard extrapolation and smoothing techniques may be used directly on the cumulative distribution values  $\hat{F}_{n+1/2}$  or probability density values  $\hat{f}_n$  to give a smoothed and extrapolated estimate of  $F(x)$  or  $f(x)$ . Similarly, given such estimated curves for a discrete subset of future times  $T$ , standard interpolation and extrapolation techniques may be used to estimate such curves for other specified values of  $T$ , or for a continuous range of  $T > 0$ .

A less direct but useful approach is to perform extrapolation and smoothing on an implied volatility function, which is then used to calculate the other functions, such as  $c(x)$ ,  $p(x)$ ,  $F(x)$ , and  $f(x)$ . The volatility rate of an asset (often simply called its volatility) is a measure of uncertainty about the returns provided by the asset. The volatility rates of a stock may typically be in the range of 0.3 to 0.5 per year.

An advantage of performing extrapolation and smoothing on implied volatility curves is that different types of volatility curves (so-called "volatility smiles") are known and can be used as a guide to the extrapolation and smoothing process to prevent "overfitting" of certain unreliable data points.

The standard method of computing implied volatilities is to invert the Black-Scholes pricing formula (see Appendix) for the actual call price  $c(x)$  or put price  $p(x)$  of an underlying asset at a given strike price  $x$ , given the underlying price  $s$  (current price of asset), risk-free rate of interest  $r$ , and and  $T$  (expiration date). When this is done for a range of values of  $x$ , an estimate of an implied volatility curve  $\sigma(x)$  is obtained. This curve may be smoothed and extrapolated by any standard method to give a smoothed curve  $\tilde{\sigma}(x)$ . Then corresponding smoothed put and call price curves may be computed using the Black-Scholes pricing formula and differentiated once or twice to give a smoothed cdf or pdf. Finally, given such estimated curves for a discrete subset of future times  $T$ , standard interpolation and extrapolation techniques may be used to estimate such curves for other specified values of  $T$ , or for a continuous range of  $T > 0$ .

Another new way to compute implied volatilities is first to compute a finite subset of cdf values  $\hat{F}_{n+1/2}$  and then to invert the Black-Scholes cdf formula (see Appendix) at these values. When this is done for a range of values of  $x$ , an estimate of a generally different implied volatility curve  $\sigma_1(x)$  is obtained, called the cdf-implied volatility curve. Again, this curve may be smoothed and extrapolated by any standard method to give a smoothed curve  $\tilde{\sigma}_1(x)$ . Then a corresponding smoothed cdf may be computed from the Black-Scholes cdf formula, and differentiated once to give a smoothed pdf. Finally, again, given such estimated curves for a discrete subset of future times  $T$ , standard interpolation and extrapolation

techniques may be used to estimate such curves for other specified values of  $T$ , or for a continuous range of  $T > 0$ .

Some advantages of using the cdf-implied volatility curve rather than the conventional implied volatility curve are that the computations are

- 5 simpler, at least from an estimate of  $F(x)$ , and that it fits better with the multivariate techniques to be discussed below.

- A particular method for finding a smoothed and extrapolated implied volatility curve  $\tilde{\sigma}_1(x, T)$  as a function of both strike price  $x$  and time  $T$  to expiration is as follows. The volatility curve is assumed to be  
 10 approximated by a quadratic formula

$$\tilde{\sigma}_1(x, T) = a_0 + a_1 x + a_2 x^2 + a_3 T + a_4 T^2 + a_5 x T, \quad (14)$$

- The coefficients  $\{a_i\}$  are determined by regression to fit the available data regarding  $\sigma_1(x, T)$  as closely as possible. Given the smoothed curve  $\tilde{\sigma}_1(x, T)$ , corresponding smoothed cdfs for different  $x$ 's and  $T$ 's) may be  
 15 computed from the Black-Scholes cdf formula for each time  $T$ , and differentiated once to give a smoothed pdf. An alternative procedure, with numerical advantages, is to use a quadratic fit like the above for a function  $\tilde{\sigma}(x, T)$ , and then invert the Black-Scholes cdf to find  $\tilde{\sigma}_1(x, T)$ . See the Appendix for the academic history of such approximations of  $\tilde{\sigma}(x, T)$ .  
 20 Another useful variation is to fit  $\tilde{\sigma}(x, T)$  with a quadratic function of  $x$  at times  $T$  which are specific expiration dates, then linearly interpolate at other times  $T$ .

#### *Treatment of multiple assets*

- The techniques described so far give probability distributions for the  
 25 future values of a single asset based on option price data for that asset. However, in many cases an investor may be concerned with multiple

assets, for example all of the stocks in his or her portfolio, or in a mutual fund, or in a certain index. Moreover, the investor may be concerned with the relations between one group of assets and another.

A general method for dealing with such questions is to generate

5 multivariate probability distributions for all assets of interest. A multivariate cdf may be written as  $F(x_1, x_2, \dots, x_n)$ , where the variables  $(x_1, x_2, \dots, x_n)$  are the values of the  $n$  assets of interest.

We will assume that we know from the techniques described above or otherwise the marginal cdfs  $F_i(x_i)$  for each of the individual variables. As  
10 a first step, we may define for each  $x_i$  a function  $y_i(x_i)$ , called a "warping function," such that  $y_i(x_i)$  is a standard normal (Gaussian) variable with mean 0 and variance 1. This is simply done by defining  $y_i(x_i)$  such that  $F_i(x_i) = N(y_i(x_i))$  for all values of  $x_i$ , where  $N(x)$  denotes the cdf of a standard normal variable. The function  $y_i(x_i)$  may be simply described in  
15 terms of  $\sigma_i(x_i)$ . See the Appendix. Under mild technical conditions such as having a marginal cdf that varies monotonically, such a warping function  $y_i(x_i)$  has a well-defined inverse warping function  $x_i(y_i)$ .

Second, we assume that we can find the historical pairwise correlations between the warped standard normal variables  $y_i(x_i)$ . These correlations  
20 may be computed by standard techniques from any available set of historical asset price data. We denote by  $C$  the  $n \times n$  correlation matrix whose entries are these historically-based correlations. Because each of the variables  $y_i(x_i)$  is standard normal, the diagonal terms of  $C$  are all equal to 1.

25 Now let  $F_C(x_1, \dots, x_n)$  denote the cdf of a multivariate Gaussian random  $n$ -tuple with zero mean and covariance matrix  $C$ . Define

$$F(x_1, x_2, \dots, x_n) = F_C(y_1(x_1), y_2(x_2), \dots, y_n(x_n))$$

Then  $F(x_1, x_2, \dots, x_n)$  is a multivariate cdf that (a) has the correct (given) marginal cdfs  $F_i(x_i)$ ; and (b) has the correct (historical) correlations between the warped standard normal variables  $y_i(x_i)$ . We use this cdf to

5 answer questions involving the variables  $(x_1, x_2, \dots, x_n)$ .

For example, the investor might have a portfolio consisting of a given quantity of each of these assets. The value of such a portfolio is the sum

$$x = h_1 x_1 + h_2 x_2 + \dots + h_n x_n, \quad (15)$$

10 where  $h_i$  represents the quantity of the  $i$ th asset in the portfolio. The investor might be interested in an estimate of the probability distribution of the value  $x$  of the whole portfolio.

Such an estimate may be obtained by Monte Carlo simulation. For such a simulation, a large number  $N$  of samples from the multivariate Gaussian cdf  $F_C(y_1, \dots, y_n)$  may be generated. Each sample  $(y_1, \dots, y_n)$  may be  
15 converted to a sample  $(x_1, x_2, \dots, x_n)$  by using the inverse warping functions  $x_i(y_i)$ . The value  $x$  of the total portfolio may then be computed for each sample. From these  $N$  values of  $x$ , the probability distribution of  $x$  (e.g., its cdf  $F(x)$ ) may be estimated.

20 In practice, it is useful to save the  $N$  multivariate samples in a large database. Then the cdf of any quantity whose value is a function of the variables  $(x_1, x_2, \dots, x_n)$  may be estimated from this database. For example, if the investor would like to know the cdf of some alternative portfolio with different quantities of each asset, this can be quickly determined from the stored database.

An investor may also determine the effect of one portfolio (or event(s) or variables such as interest rates, P/E ratios, public interest in a certain sector of the market) on another portfolio as follows. Assume that the first portfolio is represented by  $x$ , where

$$x = h_1 x_1 + h_2 x_2 + \dots + h_n x_n, \quad (30)$$

where each  $x_i$  may be viewed as the price of a portfolio component, and the second portfolio is represented by  $y$ , where

$$y = g_1 x_1 + g_2 x_2 + \dots + g_n x_n. \quad (31)$$

where each  $y_i$  may be viewed as the price of a portfolio component or more broadly as any macro-economic variable (macroeconomic, fundamental, or sector related).

Consider the "what-if" question: letting  $A$  and  $B$  be given positive constants, if  $x \geq A$  at time  $T$ , what is the probability that  $y \geq B$  at time  $T$ . This question can be answered by creating a Monte Carlo database as above for the multivariate cdf  $F(x_1, x_2, \dots, x_n)$  corresponding to time  $T$ , identifying those samples for which  $x \geq A$ , and then using only these samples to estimate the probability that  $y \geq B$ . More generally, any conditional cdf of the form  $F(x | E)$  can be estimated similarly, where  $x$  is any function of the variables  $(x_1, x_2, \dots, x_n)$  and  $E$  is any event defined in terms of the variables  $(x_1, x_2, \dots, x_n)$ .

Similarly, suppose an investor would like to know whether it is reasonable to believe that a certain stock or portfolio  $x$  will have a value greater than a given constant  $A$  at time  $T$ . This kind of question can be addressed by estimating the conditional cdf of some other related and perhaps better-understood variable (or combination of variables)  $y$  at time  $T$ , given that  $x$

$\geq A$ . If the resulting distribution for  $y$  does not look reasonable, then the investor may conclude that it is unreasonable to expect that  $x \geq A$ .

*Applications that use the probability distribution information*

5 A wide variety of techniques may be used to accumulate and process the information needed for the calculations described above and to provide the information to users directly or indirectly through third parties. Some of these techniques are described below.

As shown in figure 4, the probability distribution information can be provided to users from a host server 102 connected to a communication  
10 network 104, for example, a public network such as the Internet or a private network such as a corporate intranet or local area network (LAN). For purpose of illustration, the following discussion assumes that network 104 is the Internet.

The host server 102 includes a software suite 116, a financial database  
15 120, and a communications module 122. The communications module 122 transmits and receives data generated by the host server 102 according to the communication protocols of the network 104.

Also connected to the network are one or more of each of the following (only one is shown in each case): an individual or institutional user 108, an  
20 advertisement provider 110, a financial institution 112, a third party web server 114, a media operator 122, and a financial information provider 106.

The operator of the host server could be, for example, a financial information source, a private company, a vendor of investment services, or  
25 a consortium of companies that provides a centralized database of information.

The host server 102 runs typical operating system and web server programs that are part of the software suite 116. The web server programs allow the host server 102 to operate as a web server and generate web pages or elements of web pages, e.g., in HTML or XML code, that allow  
5 each user 108 to receive and interact with probability distribution information generated by the host server.

Software suite 116 also includes analytical software 118 that is configured to analyze data stored in the financial database 120 to generate, for example, the implied probability distribution of future prices of assets and  
10 portfolios.

The financial database 120 stores financial information collected from the financial information providers 106 and computation results generated by the analytical software 118. The financial information providers 106 is connected to the network 104 via a communication link 126 or the  
15 financial information providers may feed the information directly to the host server through a dialup or dedicated line (not shown).

Figure 4 gives a functional view of an implementation of the invention. Structurally, the host server could be implemented as one or more web servers coupled to the network, one or more applications servers running  
20 the analytical software and other applications required for the system and one or more database servers that would store the financial database and other information required for the system.

Figure 10 shows an example of a data feed 150 sent from the financial information provider 106 to the host server 102 through the  
25 communication link 126. Information is communicated to the host server in the form of messages 151, 152. Each message contains a stream of one or more records 153 each of which carries information about option prices



for an underlying asset. Each message includes header information 154 that identifies the sender and receiver, the current date 155, and an end of message indicator 158, which follows the records contained in the message.

- 
- 5 Each record 153 in the stream includes an identifier 156 (e.g., the trading symbol) of an underlying asset, an indication 158 of whether the record pertains to a put or call, the strike date 160 of the put or call, the strike price 162 of the put or call, current bid-ask prices 164 of the underlying asset, bid-ask prices 166 for the option, and transaction volumes 168
- 10 associated with the option. The financial information provider 106 may be an information broker, such as Reuters, Bridge, or Bloomberg, or any other party that has access to or can generate the information carried in the messages. The broker may provide information from sources that include, for example, the New York Stock Exchange and the Chicago Board of
- 15 Options Exchange.

- The financial database 120 stores the information received in the information feed from the financial information providers and other information, including, for example, interest rates and volatilities. The financial database also stores the results generated by the analytical
- 20 software, including probability distribution functions with respect to the underlying assets and assets that are not the subject of options.

- The probability distribution information is generated continually (and essentially in real time) from the incoming options data so that the information provided and displayed to users is current. That is, the
- 25 information is not based on old historical data but rather on current information about option prices.

In addition, other soft information can be accumulated, stored, and provided to users, including fundamental characteristics of the underlying assets, including prices, volatility values, beta, the identification of the industry to which the asset belongs, the yield, the price to book ratio, and the leverage. Other information could include calendars of earnings forecast dates, earnings forecasts, corporate action items, news items that relate to an industry, and the volume of institutional holdings.

The messages from the information provider 106 may be sent in response to requests by the host server 102, the information may be sent to the host server 102 automatically at a specified time interval, or the information may be sent as received by the information provider from its sources. The financial database 120 may be maintained on a separate server computer (not shown) that is dedicated to the collection and organization of financial data. The financial database is organized to provide logical relationships among the stored data and to make retrieval of needed information rapid and effective.

The user 108 may use, for example, a personal computer, a TV set top box, a personal digital assistant (PDA), or a portable phone to communicate with the network 104. Any of these devices may be running an Internet browser to display the graphical user interface (GUI) generated by the host server 102.

The host server 102 may provide probability distribution information on the network 104 in the form of web pages and allow the individual user 108, the financial institution 112, the third party web server 114, and the media operator 124 to view the information freely. The host company that runs the host server 102 may generate revenue by, for example, selling advertisement space on its web pages to an advertisement provider 110.

The host server 102 may also provide proprietary information and enhanced services to individual users 108, financial institutions 112, third party web servers 114, and media operators 122 for a subscription fee.

The host server 102 may have a direct link to the financial institutions 112

---

5 to provide tailored information in a format that can be readily incorporated into the databases of the financial institutions 112. Financial institutions 112 may include, for example, investment banks, stock brokerage firms, mutual fund providers, bank trust departments, investment advisers, and venture capital investment firms. These institutions may incorporate the  
10 probability distribution information generated by the analytical software 118 into the financial services that they provide to their own subscribers. The probability distribution information provided by the host server 102 enables the stock brokerage firms to provide better advice to their customers.

15 A third party web server 114 may incorporate probability distribution information into its web site. The information may be delivered in the form of an information feed to the third party host of web server 114 either through the Internet or through a dedicated or dial-up connection.

Figure 10 shows an example of a data feed 182 sent from the host server  
20 102 to the third party web server 114 through communication link 128. Data feed 182 carries messages 184 that include header information 186, identifying the sender and receiver, and records 188 that relate to specific underlying assets.

Each record 188 includes an item 190 that identifies a future date, a  
25 symbol 192 identifying the asset, risk-neutral probability density information 193 and cumulative distribution information 194. The record could also include a symbol identifying a second asset 195 with respect to

the identified future date, and so on. Other information could be provided such as a risk premium value with respect to the risk-neutral values.

Examples of third party web servers 114 are the web servers of E\*TRADE, CBS MarketWatch, Fidelity Investments, and The Wall Street

5 Journal. The third party web server 114 specifies a list of assets for which it needs probability distribution information. Host server 102 periodically gathers information from financial information provider 106 and its own financial database 120, generates the probability distribution information for the specified list of assets, and transmits the information to the third  
10 party web server 114 for incorporation into its web pages.

Examples of the media operator 124 are cable TV operators and newspaper agencies that provide financial information. For example, a cable TV channel that provides stock price quotes may also provide probability distribution information generated by the host server 102. A  
15 cable TV operator may have a database that stores the probability distributions of all the stocks that are listed on the NYSE for a number of months into the future. The host server 102 may periodically send updated information to the database of the cable TV operator. When a subscriber of the cable TV channel views the stock price quotes on a TV, the  
20 subscriber may send commands to a server computer to the cable TV operator via modem to specify a particular stock and a particular future date. In response, the server computer of the cable TV operator retrieves the probability distribution information from its database and sends the information to the subscriber via the cable network, e.g., by encoding the  
25 probability distribution information in the vertical blank interval of the TV signal.

Likewise, a newspaper agency that provides daily transaction price quotes may also provide the probabilities of stock prices rising above certain percentages of the current asset prices at a predetermined future date, e.g., 6 months. A sample listing on a newspaper may be "AMD 83 88 85

5 ▲40%"; meaning that the AMD stock has a lowest price of \$83, highest price of \$88, a closing price of \$85 that is higher than the previous closing price, and a 40% probability of rising 10% in 6 months.

The analytical software 118 may be written in any computer language such as Java, C, C++, or FORTRAN. The software may include the  
10 following modules: (1) input module for preprocessing data received from the financial data sources; (2) computation module for performing the mathematical analyses; (3) user interface module for generating a graphical interface to receive inputs from the user and to display charts and graphs of the computation results; and (4) communications interface  
15 module for handling the communications protocols required for accessing the networks.

#### *Web pages and user interfaces*

A variety of web pages and user interfaces can be used to convey the information generated by the techniques described above.

20 For example, referring to figure 5, a GUI 700 enables a user 108 to obtain a range of financial services provided by the host server 102. The user 108 may see the implied probabilities of future prices of marketable assets 706 having symbols 704 and current prices 708. The information displayed could include the probabilities 714 (or 718) of the asset prices rising above  
25 a certain specified percentage 712 (or falling below a certain specified percentage 716) of the reference price 710 within a specified period of time 720.

For the convenience of the user 108, GUI 700 includes links 730 to institutions that facilitate trading of the assets. The host company that runs the host server 102 sells advertising space 728 on the GUI 700 to obtain revenue. The GUI 700 also has links 726 to other services provided by the host server 102, including providing advice on lifetime financial management, on-line courses on topics related to trading of marketable assets, research on market conditions related to marketable assets, and management of portfolios of assets.

Referring to figure 6, the GUI 700 also may display an interactive web page to allow the user 108 to view the market's current prediction of future values of portfolios of assets. The past market price 734 and current market price 736 of the asset portfolios 732 are displayed. Also displayed is the price difference 738. The GUI 700 displays the probability 744 (or 746) that the portfolio 732 will gain (or lose) a certain percentage 740 within a specified time period 742. Examples of portfolios include stock portfolios, retirement 401K plans, and individual retirement accounts. Links 748 are provided to allow the user 108 to view the market's current forecast of future price trends of the individual assets within each portfolio.

Referring to figure 7, in another user interface, the GUI 700 displays an interactive web page that includes detailed analyses of past price history and the market's current forecast of the probability distribution of the future values of a marketable asset over a specified period of time. The GUI 700 includes price-spread displays 750 representing the cumulative distribution values of the predicted future prices of an asset over periods of time. The price-spread display 750a shows the price distribution data that was generated at a time three months earlier. A three-month history of the actual asset prices is shown as a line graph for comparison to give the user

108 a measure of the merit of the price distribution information. The price-spread display 750b represents the predicted cumulative distribution values of the asset prices over a period of one month into the future. The left edge of display 750b, of course, begins at the actual price of the asset as of the end of the prior three-month period, e.g., the current DELL stock price of \$50. The probability distribution information implies, for example, a 1% probability that the stock price will fall below \$35, and a 99% probability that the stock price will fall below \$80 in one month. GUI 700 includes table 752 that shows highlights of asset information and graph 754 that shows sector risks of the asset. A box 755 permits a user to enter a target price and table 757 presents the probability of that price at four different future times, based on the calculated implied probability distributions.

Referring to figure 8, in another approach, a window 402 is displayed on a user's screen showing financial information along with two other windows 408 and 410 showing probability distribution information. The individual user 108 could have previously downloaded a client program from the host server 102. When the user is viewing any document, e.g., any web page (whether of the host server 102 or of another host's server), the user may highlight a stock symbol 404 using a pointer 406 and type a predetermined keystroke (e.g., "ALT-SHIFT-Q") to invoke the client program. The client program then sends the stock symbol as highlighted by the user to the host server 102. The host server 102 sends probability distribution information back to the client program, which in turn displays the information in separate windows 408 and 410.

When the client program is invoked, a window 422 may be displayed showing the different types of price information that can be displayed. In the example shown, the "Probability distribution curve" and "Upper/lower

estimate curves" are selected. Window 408 shows the price range of AMD stock above and below a strike price of \$140 from July to December, with 90% probability that the stock price will fall between the upper and lower estimate curves. Window 410 shows the probability density curve  $f(x)$  for AMD stock for a future date of 8/15/2000. The user may also specify a default function curve, such that whenever an asset name is highlighted, the default function curve is displayed without any further instruction from the user.

10 Tabular data such as those shown in TABLE 1 may be generated by the host server 102 and transmitted over the network 104 to devices that have limited capability for displaying graphical data. As an example, the individual user 108 may wish to access asset probability distribution information using a portable phone. The user enters commands using the phone keypad to specify a stock, a price, and a future date. In response, the  
15 host server 102 returns the probability of the stock reaching the specified price at the specified future date in tabular format suitable for display on the portable phone screen.

Referring to figure 9, a portable phone 500 includes a display screen 502, numeric keys 506, and scrolling keys 504. A user may enter commands  
20 using the numeric keys 506. Price information received from the host server 102 is displayed on the display screen 502. Tabular data typically includes a long list of numbers, and the user may use the scroll keys 504 to view different portions of the tabular data.

In the example shown in display screen 502, the AMD stock has a current  
25 price of \$82. The cumulative distribution values  $F(x)$  for various future prices on 8/15/2000 are listed. The distribution indicates a 40% probability that the stock price will be below \$80 implying a 60% probability of the



stock price being above \$80. Likewise, the distribution indicates an 80% probability that the stock price will be at least \$90, implying a probability of 20% of the stock price being above \$90.

The visualization techniques discussed below are useful in enabling users

5 to visualize and quickly understand information that relates to assets.

*Visualization of implied probability distributions of future prices*

As shown in figure 11, a visualization device 10 displays cumulative probability distribution values of predicted relative future prices of Dell Computer Corporation stock with respect to a current date 12 of July 1, 2000. The price 14 on July 1, 2000, is shown as being \$41 lower than the price 16 on February 1, 2000, which itself is set at an arbitrary starting value of \$0 for purposes of display. The display could be provide in actual price terms, as a price change, or in terms of percentage return. The probability distribution data on which the visualization device 10 is based 15 may be generated by, for example, the method discussed in the parent patent application.

The predicted cumulative distribution values of the prices of Dell stock over a period of several months into the future are illustrated by an envelope 16 that begins at a point 18 and opens to the right

20 The envelope 16 is divided into stripes 22, 24, 26, 28, 30, each of which also begins at point 18 and opens to the right. Stripe 22, for example, indicates a range of prices (all of which are below the current price) at each date in the future and indicates the predicted odds (10%) that the price will fall within that stripe. Similarly, stripe 26 indicates a range of 25 prices (above and below the current price) with an expected 40% odds of occurring on various dates in the future. The odds of falling either above

or below envelope 16 are, as indicated, less than 1%. Each stripe is displayed in a different color, and the colors are chosen to permit a viewer to visualize the different stripes easily.

A similar envelope 32 starts at the nominal \$0 price on February 1, 2000,

5 and ends on the current date. Envelope 32 represents the cumulative distribution values of the prices of Dell stock that were predicted as of February 1, 2001. The actual price history of Dell stock between February 1, 2000, and the current date is illustrated by the line 34. The extent to which the actual price history of line 34 matches the predicted cumulative  
10 distribution values gives a visual indicator to the user of the validity of the prediction approach.

The combination of color, text, and data illustrated in figure 11 enable an investor to assess the performance of an asset over time relative to his price expectations.

15 The visualization device of figure 11 is also useful for assets other than stocks, including mutual funds, and for portfolios of assets.

Figure 12 presents information similar to figure 11, but is expressed with respect to projected return percentage rather than price. The example shown in figure 12 relates to Check Point Software Technologies Ltd  
20 stock as of a current date 66 of October 24, 2000. The x-axis represents return percentage with respect to a start date. Line 62 shows the historical return with respect to the stock price on the start date of January 1, 2000 at point 67. On the current date 66, the cumulative return on the price of the stock since start point 67 is approximately 200%.

25 An envelope 68 starts at point 66 and opens to the right. The envelope 68 illustrates the projected odds of the percentage return being within certain

ranges on each day for several months into the future relative to the original start point 66. The ranges are expressed as stripes 52, 54, 56, 58, and 60. The envelope and stripes are centered on a trend line 50 that has a slightly positive slope to reflect the probability of future price levels

5 generated by a mathematical algorithm that is based on the implied volatility of the options market. The algorithm is described in the related pending United States patent application 09/641,589, filed 08/18/2000.

For example, the projected odds that the return (relative to the start point 67) will be between 50 and 100% on May 1, 2000, is 10%.

10 The same kind of data used to generate the display of figure 11 is used to generate the device of figure 12 except that the data is processed to convert the price data into change of price data for plotting along the x-axis.

Figure 13 is similar to figure 12, except that the effect of the occurrence of  
15 the long-term capital gain tax rate transition (identified as the vertical line 80 that is one year after the start date 82). After the date represented by line 80, any sale of the stock would produce a lower tax impact and a higher effective return, than under the assumption of short-term capital gain tax rate, prior to that date. For that reason, the envelope 84 is shifted  
20 upward and exploded for periods after the transition date.

#### *Visualization of asset style*

Figure 14 shows another visualization device that reflects an asset fund style analysis that evaluates an asset fund (e.g., a mutual fund) by  
25 comparing its historical returns to those achieved by a set of basic asset classes (e.g., cash, bonds, large-cap growth stocks, large-cap value stocks).

The first step of the style analysis is a one-time selection of basic asset classes, which should be mutually exclusive and exhaustive, to represent all asset types of interest. In one example of classes (listed below) there are seventeen market indices, seven of which represent stocks and the remainder of which represent bonds.

The second step of style analysis determines the exposure of a given mutual fund to these indices. This is achieved by solving an asset class factor model, in which a fund return is expressed as a linear combination of returns from basic asset classes plus a residual. The exposures are determined by minimizing the variance of residuals using one-year weekly data. It is believed that one-year weekly data can reflect a fund style more accurately. In addition, fund exposures to basic asset classes are constrained to be non-negative and to sum to one.

The third step of style analysis is to present the results in a form that provides meaningful investment information. Style analysis results for a given fund consist of percentages in each basic asset class, with the dominant percentages determining the fund's style. Style drift for a given fund is based on determining style changes over the most recent five years.

In figure 14, the results of the analysis are displayed. The colors of the respective cells 102 indicate how much of the fund's performance is explained by regression to the style associated with the row in which the cell appears, during the period represented by the column in which the cell appears.

The example shown in figure 14 identifies each of seventeen indices (styles 100) that are of interest to a broad group of individual investors. For example, the style LG refers to a set of stocks that are characterized as

Large Capitalization Growth. The full list of groups in this example follows:

- |    |     |                                    |
|----|-----|------------------------------------|
|    | 1.  | Large-Cap Growth (LG)              |
|    | 2.  | Large-Cap Value (LV)               |
| 5  | 3.  | Mid-Cap (MC)                       |
|    | 4.  | Small-Cap (SC)                     |
|    | 5.  | European Stocks (EU)               |
|    | 6.  | Japanese Stocks (JP)               |
|    | 7.  | Emerging Markets (EM)              |
| 10 | 8.  | Cash (TB)                          |
|    | 9.  | Intermediate Government Bonds (GI) |
|    | 10. | Long-term Government Bonds (GL)    |
|    | 11. | Intermediate Corporate Bonds (CI)  |
|    | 12. | Long-term Corporate Bonds (CL)     |
| 15 | 13. | Corp Junk Bonds (HY)               |
|    | 14. | Mortgages-Backed Securities (MG)   |
|    | 15. | Real Estate (RE)                   |
|    | 16. | Municipal Bonds (MU)               |
|    | 17. | Global Bonds (GG)                  |

Thus, for cell 102, the regression indicates that about 45% of the fund's performance is correlated to the LG style for that period in 2000.

The values determined by the regression are displayed in a grid with style on the vertical axis and time on the horizontal axis. The color of each cell

5      102 indicates the percentage in accordance with the percentage scale shown on the right.

The resulting visualization device enables an investor to assess the performance of the asset over time relative to his investment preferences and strategy.

10      *Visualization of recent market activity*

The ability to track the activity of a market of assets (such as stocks or mutual funds) as the activity unfolds is of great interest to investors. Many investors rely on daily publications of tabular data that presents information such as volume, price change, asset identification, and performance.

15      The visualization device shown in figure 15 collects, condenses, and enhances such information in a way that improves the ability of an investor to visually and quickly grasp recent and current market activity.

20      The displays are updated continually and quickly throughout a trading day.

As shown in figure 15, the visualization device 120 includes a radar-like display that is divided into sectors 122 arranged around a central point 124. The device is also divided into rings 126 that are centered on point 124 and filled with different colors to distinguish the different rings visually.

25

Each of the sectors 122 is associated with an industry or sector of interest to investors, for example the technology sector or the financial sector. The size of each sector depends on the proportion of the asset items being displayed for the sector relative to the total number of asset items being depicted for the whole universe.

Each of the rings represents a different percentage of price change during a recent period (e.g., during a single trading day). The rings are arranged with the largest percentage decline near the middle of the radar and the largest percentage increase near the periphery.

Within each of the sectors, small dots 128 are displayed each representing a selected stock or asset within the industry sector represented by the radar sector. The distance of each dot from the central point 124 represents the percentage price change of the corresponding stock at a given time during a trading day. Gray dots represent small capitalization stocks; black dots represent large capitalization stocks.

When multiple stocks in a sector have the same percentage change (e.g., at location 130), the dots are displayed at different angular positions relative to the central point, to convey to the viewer an impression of the distribution of the percentage changes within each sector.

#### *Implementation details*

The visualization elements described above can be displayed on a wide range of devices, including desktop and laptop computers, personal digital assistants, portable telephones, publicly viewed large-screen displays, or closed circuit or broadcast/cable television monitors.

The visualization elements can be displayed alone or embedded in other displayed material, including other financial information, general news

information, or program material. For example, the elements can be displayed as part of a website page dedicated to financial information or as part of a general web portal page. The elements can be displayed as part of a broadcast or cable TV program.

- 
- 5     The raw data from which the visualization elements are created may be obtained on-the-fly electronically and/or may be stored as needed either locally or centrally. Software that processes the raw data to generate the derived values to be represented in the visualization elements may run locally or may be run remotely (and then downloaded to a local display).
- 10    Software that processes the derived values to produce the visualization elements may be handled similarly.

The raw data, the derived values, and the visualization elements can be updated more or less frequently, though in many cases real-time updates are especially useful.

- 15    Each of the visualization elements could be made interactive by enabling a user to provide inputs, for example, mouse clicks, that indicate how the user wishes to alter the manner in which the elements are displayed, or the selection of data contained in them. Configuration features can be provided to enable the user to configure what information he receives, in
- 20    what form it is displayed, and how often and how currently he receives it.

Other implementations are within the scope of the following claims.

- For example, with respect to the visualization element shown in figure 15, the overall shape of the element could be other than round, the sectors could be other than simple pie shapes, the rings could be other than simple
- 25    rings, the individual dots could be replaced by other icons, the dots or other icons could be arrayed in other arrangements from the center, and



visible features other than color could be used to distinguish different portions of the display.

A wide range of variants is also possible with respect to the visualization elements shown in the figures.

---

## Appendix A

Three aspects of the invention are:

1. The recognition of the desirability of displaying to a financial investment customer in real time, for example on a World Wide Web site, the probability distribution governing the price of a particular asset (*e.g.*, a stock) at a selected future time.
2. The recognition that such probability distributions can be derived from option prices for that asset, or for related assets, which are readily available in real time.
3. The recognition that probability distributions involving several asset prices simultaneously are useful to investment customers in several contexts, especially in exploring hypothetical scenarios, and that single asset distributions such as (but not restricted to) the above can be meaningfully incorporated into multivariate distributions, manageably determined.

In this appendix we first describe a basic method for deriving probability distributions for single assets from option prices. We next describe improvements on this basic method to address various practical issues. Then we take up the multivariate case and show how to extend this kind of single asset price distribution, or any other, to the multivariate case. Finally, we consider a number of novel multivariate applications, with emphasis on scenario exploration.

### 1 Basic method

A call option is an option to buy an asset (*e.g.*, a stock) at a certain price  $x$  (called the strike price) on a given expiration date  $T$  days in the future. (An option exercisable only on the expiration date is called a European-style option; for simplicity we will consider in this discussion only this type of option.<sup>1</sup>) Similarly, a put option is an option to sell an asset at a strike price  $x$  on a given expiration date. (The "European-style" assumption of no possible early exercise is more important here, but can also be ignored for puts that are not too deeply "in the money.")

Let  $c(x)$  denote the price of a call option on an asset at strike price  $x$ , and  $p(x)$  the price of a put option. Such prices are established by options market-makers. We have realized that such prices implicitly contain information about a "market view" of the probability distribution of the price of that asset at the expiration date.

In a simple but precise form, this market view can be stated as follows. Suppose that we were given the call price curve  $c(x)$  or the put price curve  $p(x)$  as a continuous function of the strike price  $x$  for all  $x > 0$ . Then, the second derivative of either the call or the put price curve is the market view of the risk-neutral probability density function (pdf)  $f(x)$  of the asset price at the expiration date. In other words,  $f(x) = c''(x) = p''(x)$ .

The idea that option prices determine some kind of implied probability distribution is fairly well known in the financial literature. The idea that a pdf can be computed by taking the second derivative of a continuous option price curve is known in the academic literature, but it does not appear to be very well known. For example, the standard textbook "Options, Futures,

<sup>1</sup> Even allowing for possible early exercise, most liquidly traded call options without large dividends can be treated as if there were no possibility of such exercise, since sale of the option is usually a better alternative; therefore, these call options behave similarly to European-style options.

and Other Derivatives," by John C. Hull (Fourth Edition, 1999; Prentice-Hall) mentions implied probabilities, but not the second-derivative method. The best reference that we have been able to find is J. C. Jackwerth and M. Rubinstein, "Recovering probability distributions from option prices," *J. Finance*, vol. 51, pp. 1611-1631 (1996), which has only six prior references.

The risk-neutral distribution (at a fixed future time  $T$ , for a fixed asset) is defined as the price distribution that would hold if market participants were neutral to risk, which they generally are not. However, many asset pricing theories, such as those underlying Black-Scholes option theory and most of the variations found in the Hull book above, allow for the true risk-averse asset price distribution to be obtained from the risk-neutral distribution  $f(x)$  just by adjusting the latter by an appropriate risk premium: If there are no dividends, the true distribution is just  $f(xe^{(\mu-r)T})$ , where  $\mu - r$  is the expected annual return rate for the stock in excess of the risk free rate  $r$ . We use a variation on this simple format, slightly modified to allow for dividends (see below), though our invention could also work well with a more complicated adjustment. In this format, a value for  $\mu - r$  must still be supplied. We use as a default the "consensus estimate" taken from the textbook "Active Portfolio Management" (1995) by Grinold and Kahn. These authors note a long-term average value of the risk premium to be 6% per year, and suggest multiplying this number by the stock's beta to get  $\mu - r$ . The parameter beta is the slope of the line giving a regression of the stock in question against a market portfolio, often taken as the S&P 500. This is the well-known CAPM estimate for the expected excess return. Whether good or bad, its stature as a consensus estimate makes it suited to our aim of providing a market view, though it is only a default. Our invention, which provides the risk-neutral component of the probabilities, could work with other estimates for the risk-averse adjustment parameter  $\mu - r$  and with any explicit scheme for adjusting the risk neutral probability density to the risk-averse probability density. It is worth pointing out that, for shorter time periods—even a month or two—the risk adjustment required is small and generally overwhelmed by fluctuations in the risk-neutral distribution itself.

We give a brief proof that the second derivative procedure gives the correct risk-neutral probability distribution. As in Hull, we may calculate the European call or put price as an expected value in the risk-neutral distribution.

If the actual value of the asset on the expiration date is  $v$ , then the value of a call option at strike price  $x$  is  $\max\{v - x, 0\}$ , and the value of a put option is  $\max\{x - v, 0\}$ . If the actual value is a random variable with pdf  $f(v)$ , then the expected value of a call option at  $x$  at the expiration date is

$$c_T(x) = E_v[\max\{v - x, 0\}] = \int_x^\infty (v - x)f(v) dv,$$

and the expected value of a put option at  $x$  at the expiration date is

$$p_T(x) = E_v[\max\{x - v, 0\}] = \int_0^x (x - v)f(v) dv.$$

The current values  $c(x)$  and  $p(x)$  may be obtained by discounting  $c_T(x)$  and  $p_T(x)$  by  $e^{-rT}$ , where  $r$  is the risk-free interest rate, but for our purposes, forecasting probability distributions at time  $T$ , we do no discounting, and henceforth just write  $c(x) = c_T(x)$ ,  $p(x) = p_T(x)$ .

Parenthetically, from these expressions we observe that

$$p(x) - c(x) = \int_0^\infty (x - v)f(v) dv = x - E_v[v] = x - s^*,$$

where  $s^* = E_v[v]$  is the expected value of the asset at the expiration date under the risk-neutral distribution. (If there are no dividends, then  $s^* = se^{rT}$ ; if there are dividends, then in general it is necessary to subtract from  $se^{rT}$  the value at time  $T$  of the dividends.) This well-known relation is called put-call parity; it shows why either price curve carries the same information.

From the above expression for  $c(x)$ , it follows that its first derivative is

$$c'(x) = - \int_x^\infty f(x) dx = F(x) - 1,$$

where  $F(x) = \int_0^x f(v) dv$  is the cumulative distribution function (cdf) of the random variable  $v$ . To prove this, note that  $v - x = \int_x^v dz$ . Therefore

$$c(x) = \int_x^\infty (v - x) f(v) dv = \int_x^\infty dv \int_x^v dz f(v) = \int_x^\infty dz \int_z^\infty dv f(v) = \int_x^\infty dz (1 - F(z)),$$

where we interchange the variables  $v, z$  to integrate over the two-dimensional region  $\mathcal{R} = \{(v, z) : x \leq z \leq v\}$ . The last expression implies that  $c'(x) = -(1 - F(x))$ .

From put-call parity, it follows similarly that

$$p'(x) = 1 + c'(x) = F(x).$$

Since the cdf and pdf are related by  $F'(x) = f(x)$ , these expressions in turn imply that the second derivative of either  $c(x)$  or  $p(x)$  is the pdf  $f(x)$ :

$$c''(x) = p''(x) = F'(x) = f(x).$$

The general character of the option price curves  $c(x)$  and  $p(x)$  is therefore as follows:

- For all  $x$  less than the minimum possible value of  $v$  (i.e., such that  $F(x) = 0$ ),  $c(x) = E_v[v] - x = s^* - x$  and  $p(x) = 0$ . In other words,  $c(x)$  is a straight line of slope  $-1$  starting at  $c(0) = E_v[v] = s^*$ , while  $p(x) = 0$ .
- For all  $x$  greater than the maximum possible value of  $v$  (i.e., such that  $F(x) = 1$ ),  $c(x) = 0$  and  $p(x) = x - s^*$ . In other words,  $p(x)$  is a straight line of slope  $+1$  and  $x$ -intercept  $s^*$ , while  $c(x) = 0$ .
- These two line segments are joined by a continuous convex  $\cup$  curve whose slope increases from  $-1$  to  $0$  for  $c(x)$ , and from  $0$  to  $+1$  for  $p(x)$ .

We note that the fact that the mean  $E_v[v]$  of the pdf  $f(x)$  is  $s^*$ , the value in future dollars at time  $T$  of the underlying price  $s$  (less the value of any dividends), implies that option prices must be constantly adjusted to reflect changes in the underlying price  $s$ , even if there is no market activity in the options.

The fact that  $s^* = E_v[v]$  also implies that an option price curve can make no prediction about the general direction of the underlying price  $s$ . However, the option price curve does predict the shape of the pdf  $f(x)$ , and in particular its volatility.

### 1.1 Approximations based on a finite subset of bid-asked option prices

In practice, option prices  $c(x)$  and  $p(x)$  are quoted only for a finite subset of equally-spaced strike prices  $x$ , namely  $x = n\Delta$  for integer  $n$  and spacing  $\Delta$ . We denote  $c(n\Delta)$  and  $p(n\Delta)$  by  $c_n$  and  $p_n$ , respectively. Moreover, quotes specify only a bid-asked spread, not exact prices. In this subsection we give methods for dealing with these problems. (Most of the Jackwerth-Rubinstein paper (*op. cit.*) is concerned with these kinds of curve-fitting problems.)

The first derivatives  $c'(x)$  and  $p'(x)$  at  $x = (n + \frac{1}{2})\Delta$  may be estimated by the first differences

$$\hat{c}_{n+\frac{1}{2}} = \frac{c_{n+1} - c_n}{\Delta}; \quad \hat{p}_{n+\frac{1}{2}} = \frac{p_{n+1} - p_n}{\Delta}.$$

The corresponding estimates of the cdf  $\hat{F}_{n+\frac{1}{2}} = F((n + \frac{1}{2})\Delta)$  are

$$\hat{F}_{n+\frac{1}{2}} = 1 + \hat{c}_{n+\frac{1}{2}}; \quad \hat{F}_{n+\frac{1}{2}} = \hat{p}_{n+\frac{1}{2}}.$$

Thus, using both bid and ask prices for both European-style puts and calls, one can compute four different estimates for the cdf  $\hat{F}_{n+\frac{1}{2}}$ , which can then be combined into a single estimate. This combination will preferably take into account whether  $x = (n + \frac{1}{2})\Delta$  is much less than the underlying price  $s$  ("deep out-of-the-money"), near  $s$  ("near the money"), or much greater than  $s$  ("deep in-the-money"), according to the different patterns of setting bid-asked spreads in these different ranges. Another consideration is avoiding quotes near prices where early exercise is likely, such as deep in-the-money puts.

Similarly, the second derivatives  $c''(x)$  and  $p''(x)$  at  $x = n\Delta$  may be estimated by the first differences of the estimates of the first derivatives; e.g.,

$$\hat{c}''_n = \frac{\hat{c}_{n+\frac{1}{2}} - \hat{c}_{n-\frac{1}{2}}}{\Delta} = \frac{c_{n+1} - 2c_n + c_{n-1}}{\Delta^2}.$$

We may take  $\hat{c}''_n$  or  $\hat{p}''_n$  or some combination as above as our estimate  $\hat{f}_n$  of the pdf  $f(n\Delta)$ .

Note that since  $f(x) \geq 0$ , option prices should satisfy a convexity condition, e.g.,  $c_{n+1} - 2c_n + c_{n-1} \geq 0$  for call option prices. Indeed, violation of this condition would allow making money via a risk-free "butterfly straddle" involving buying one call option at  $(n+1)\Delta$  and another at  $(n-1)\Delta$ , and selling two call options at  $n\Delta$ . A similar result holds for put options.

### 1.2 Dynamic estimates

The methods considered in the previous subsection allow estimation of the cdf and pdf at a subset of  $\Delta$ -spaced values of  $x$ , based on a static set of option quotes at a particular time.

As previously noted, however, option prices must change continually in response to changes in the underlying price  $s$ . Let  $s^*$  denote the corresponding forward price at expiration (the price  $s$  evaluated with interest). Suppose this price (measured in dollars at expiration) moves up (or down) by a small amount, an increment  $\epsilon$  in its logarithm, say, with little or no change in volatility. Here  $\epsilon$  may be viewed as, approximately, the percentage move  $\delta/s^*$  caused by a move of  $\delta$  in the (forward) stock price. We expect in this situation that (forward) probability distribution for the stock price will just be shifted by  $\epsilon$  in the log domain. That is, the distribution will appear to be identical there, except with a mean shifted by  $\epsilon$ . Thus, the value of the new cdf at  $x = e^{\ln x}$  is  $F(e^{(\ln x - \epsilon)}) = F(x/a)$ , where  $F$  denotes the original cdf with distribution mean

$s^*$ , and  $a = e^\epsilon$ . A reasonable call price functional equation that gives the same effect, upon differentiation, is

$$ac(s^*, x/a) = c(as^*, x),$$

where  $c(s^*, x)$  denotes the price, in dollars at expiration, for a call option at strike  $x$  when the underlying is at price  $s$ . Note in this equation that all other variables, such as volatility, are assumed to be the same, which will only be approximately true, even for very small values of  $\epsilon$ .

But, assuming this approximation, we can think of an option price at strike  $x$ , measured when the (forward) price has moved to  $as^*$ , for  $a$  near 1, as giving instead  $a$  times the price of an option at strike  $x/a$ , but corresponding to the current underlying price  $s^*$ . Considering all the strikes at which options are frequently quoted, and thinking additively, we can effectively observe  $c(x)$  (and  $p(x)$ ) for a different subset of approximately equally-spaced strike prices, roughly  $x = n\Delta - \delta$  for various values of  $\delta = \epsilon s^*$ . Some care must, of course, be taken to ensure simultaneity of prices, of option and underlying. For this reason, we may prefer to consider the values of  $n\Delta$  (corresponding to the various standard strike values) separately, and synchronize observed time of sales for an option at a given strike with the underlying security. Implied volatilities (discussed below) could be monitored, to ensure their changes relative to  $\epsilon$  were small.

Using a similar technique to that described in the above paragraphs, meaningful average option prices for a given strike can also be computed, using thin strike intervals and using either short time intervals or time series methods (time averages weighting the present more than the past). Note that, without the framework described in this subsection, the computation of "average" option prices at a given strike are problematic when the stock price varies in the period over which the average is taken.

To summarize: Given enough movements of the underlying price, we can effectively observe prices and compute estimates as above for a much more finely quantized subset of strike prices  $x$ , and provide a framework for improving accuracy through averaging methods.

## 2 Methods for extrapolation and smoothing

There are two major limitations to the basic methods of the previous section. One is that option quotes are available only for certain expiration dates. Another, not so obvious, is that option quotes are reliable primarily for options in which there is substantial market activity. These would typically be nearer-term options at strike prices near the money (the underlying price).

To extend our prediction methods to times other than expiration dates and over wider ranges of strike prices (and also to help reduce "noise" in our displays), we use extrapolation and smoothing techniques. We have found that it is advantageous to do extrapolation and smoothing in the volatility domain.

There are many reasons for this advantage. For example, option practitioners are well aware of the kinds of shapes that the volatility curves (sometimes called "volatility smiles") have had historically, in various markets, and how these curves vary with time; this can be a guide to imposing structure on the smoothing curves to prevent overfitting of possible artifacts. Many records have been kept of the volatilities implied by option prices, and it is easy to examine how in the past they have changed with respect to price behavior. For example, the Chicago Board of Options Exchange makes public its average near-the-money volatility index (now called VIX) for S&P 100 options back to 1986. Finally, it is easier to work visually with volatility curves, which

would theoretically be flat if  $f(x)$  were lognormal, than with visual differences in near-lognormal pdfs, which can all look very much alike. Mathematically, model improvements can be made in the volatility domain just by changing coefficients of low-degree polynomial approximations, even though these affect higher-order terms in power series for the corresponding cdfs or pdfs.

The following subsections explain more precisely how to work in the volatility domain.

## 2.1 Lognormal pdfs

The standard Black-Scholes theory of option pricing (see Hull, *op. cit.*) yields a lognormal pdf  $f(v)$  whose expected value is  $E_v[v] = s^*$ , such that  $\ln v$  is a Gaussian (normal) random variable with variance  $\sigma^2 T$ , where the parameter  $\sigma$  is called the volatility rate of the asset, and  $T$  is the time to expiration. By a standard property of lognormal distributions, this implies that the mean of  $\ln v$  is  $E_v[\ln v] = \ln s^* - \sigma^2 T/2$ .

From this pdf follows the famous Black-Scholes call option pricing formula [Hull, Appx. 11A]:

$$c(x) = E_v[\max\{v - x, 0\}] = s^* N(d_1(x)) - x N(d_2(x)),$$

where  $N(d_1(x))$  and  $N(d_2(x))$  are values of the cumulative distribution function of a Gaussian random variable of mean zero and variance 1 at the points

$$\begin{aligned} d_1(x) &= \frac{\ln(s^*/x) + \sigma^2 T/2}{\sigma\sqrt{T}} = d_2(x) + \sigma\sqrt{T}; \\ d_2(x) &= \frac{\ln(s^*/x) - \sigma^2 T/2}{\sigma\sqrt{T}} = \frac{E_v[\ln v] - \ln x}{\sigma\sqrt{T}}. \end{aligned}$$

(Recall that our version of the call price is not discounted, and is given in dollars at time  $T$ , and that  $s^*$  is today's stock price, valued in dollars at time  $T$ , less the value of any dividends.) Note that  $\sigma\sqrt{T}$  is the standard deviation of  $\ln v$ ; therefore  $-d_2(x)$  is just  $\ln x$ , measured in standard deviations from the mean  $E_v[\ln v]$ .

Similarly, by put-call parity, we have the Black-Scholes put option pricing formula

$$p(x) = c(x) + x - s^* = s^*(N(d_1(x)) - 1) - x(N(d_2(x)) - 1) = xN(-d_2(x)) - s^*N(-d_1(x)).$$

Taking the derivative with respect to  $x$ , and using  $s^*N'(d_1(x)) = xN'(d_2(x))$  and  $d'_1(x) = d'_2(x)$  (the latter equation holding under the assumption of constant volatility, which we will later drop), we obtain

$$F(x) = c'(x) + 1 = -N(d_2(x)) + 1 = N(-d_2(x)).$$

Now  $F(x)$  is the probability that  $v \leq x$ , which is equal to the probability that  $\ln v \leq \ln x$ , which since  $\ln v$  is Gaussian with mean  $E_v[\ln v]$  and standard deviation  $\sigma\sqrt{T}$  is given by

$$F(x) = \Pr\{v \leq x\} = \Pr\{\ln v \leq \ln x\} = N\left(\frac{\ln x - E_v[\ln v]}{\sigma\sqrt{T}}\right) = N(-d_2(x)).$$

Thus we have verified that the Black-Scholes pricing formulas give the correct cdf  $F(x)$ . The derivative of  $F(x)$  will thus yield the correct lognormal distribution  $f(x) = F'(x)$ .

## 2.2 Characterization of general cdfs

Now let  $F(x)$  be an arbitrary cdf on  $\mathbb{R}_+$ ; i.e., a function that monotonically increases from 0 to 1 as  $x$  goes from 0 to infinity. For simplicity we will assume that  $F(x)$  is strictly monotonically increasing; i.e.,  $f(x) = F'(x) > 0$  everywhere. Then there exists a continuous one-to-one "warping function"  $y : \mathbb{R}_+ \rightarrow \mathbb{R}$  such that  $F(x) = N(y(x))$  everywhere; i.e., such that the probability that a random variable  $v$  with cdf  $F(x)$  will satisfy  $v \leq x$  is equal to the probability that a standard Gaussian random variable  $n$  with mean zero and variance 1 will satisfy  $n \leq y(x)$ . Similarly, there is an inverse warping function  $x(y)$  such that  $F(x(y)) = N(y)$ .

Given the warping function  $y(x)$ , the cdf  $F(x)$  may be retrieved from the relation  $F(x) = N(y(x))$ . Therefore the cdf  $F(x)$  completely specifies the warping function  $y(x)$ , and *vice versa*; i.e., both curves carry the same information.

If  $F(x)$  is the cdf of a lognormal variable  $v$  such that  $\ln v$  has mean  $E_v[\ln v] = \ln s^* - \sigma^2 T/2$  and variance  $\sigma_v^2 = \sigma^2 T$ , as in the previous subsection, then the warping function is given by

$$y(x) = -d_2(x) = \frac{\ln x - (\ln s^* - \sigma^2 T/2)}{\sigma \sqrt{T}} = \frac{\ln x - E_v[\ln v]}{\sigma_v}.$$

For this reason we may sometimes write  $y(x)$  as  $-d_2(x)$ , even when the cdf is not lognormal so that the right-hand equation above for  $d_2(x)$  does not hold.

## 2.3 Implied volatilities

If  $f(x)$  is not lognormal, then the Black-Scholes pricing formulas do not hold. Nonetheless, given an option price  $c(x)$  or  $p(x)$ , it is common practice to define the implied volatility  $\sigma(x)$  as the value of  $\sigma$  such that the Black-Scholes pricing formula holds, for a given  $x$ ,  $s$  and  $T$ .

The implied volatility curve  $\sigma(x)$  so defined is a function of the strike price  $x$ , which is constant if and only if the pdf  $f(x)$  is actually lognormal. In practice, it is typically a convex U curve, called a "volatility smile." See, e.g., Hull, Chapter 17.

From Subsection 2.1, we can see that there is a second method of calculating implied volatilities, as follows. Suppose that we have an estimate of the cdf  $F(x)$ . Define the cdf-implied volatility  $\sigma_1(x)$  as the value of  $\sigma$  such that the Black-Scholes cdf formula  $F(x) = N(-d_2(x, \sigma, T))$  holds, for a given  $x$ ,  $s$  and  $T$ .

The first method has the advantages of being defined directly from raw price data, and of being well understood in the financial community. However, the second method has the following advantages:

1. It is easier to calculate, at least from estimates of  $F(x)$ ;
2. It gives a simpler and arguably more intuitive relationship between volatility and the cdf  $F(x)$ . If we use the traditional implied volatility  $\sigma(x)$ , then the relationship is instead

$$F(x) = N(-d_2(x)) + \frac{\partial c}{\partial \sigma} \sigma'(x).$$

3. It fits better with the multivariate theory to be developed below.



We have observed that the two curves  $\sigma(x)$  and  $\sigma_1(x)$  seem to be fairly similar, at least as to the direction of their slope, and are generally not too far apart in value "near the money". Also  $\sigma_1(x) = \sigma(x)$  whenever  $\sigma(x)$  has zero slope, though  $\sigma_1(x)$  is a little smaller than  $\sigma(x)$  when the slope  $\sigma'(x)$  is negative (which often occurs for stocks). See the above equation. Finally, one function is as *ad hoc* as the other. Therefore, because of the above reasons, we generally prefer to use the cdf-implied volatility curve  $\sigma_1(x)$ .

In any case, it is clear that either  $\sigma(x)$  or  $\sigma_1(x)$  contains the same information as any of the curves  $c(x)$ ,  $p(x)$ ,  $F(x)$  or  $f(x)$ . From  $\sigma(x)$  or  $\sigma_1(x)$  we can recover  $c(x)$  or  $F(x)$  using the Black-Scholes call option pricing or cdf formula, and from this we can obtain all other curves.

## 2.4 Extrapolation and smoothing in the volatility domain

The volatility curve  $\sigma(x)$  or  $\sigma_1(x)$  may be calculated pointwise from the corresponding curve  $c(x)$  or  $p(x)$  to give a set of values at a finite subset of strike prices  $x$ . Each of these values may be deemed to have a certain degree of reliability.

It is then a standard problem to fit a smoothed and extrapolated curve  $\bar{\sigma}(x)$  or  $\bar{\sigma}_1(x)$  to these points, taking into account their relative reliabilities. Any standard smoothing and extrapolation method may be used. In general, the usual problems of avoiding overfitting or oversmoothing must be addressed.

It is well-known that implied volatilities also vary with time. We generally wish to estimate curves  $\bar{\sigma}(x, T)$  or  $\bar{\sigma}_1(x, T)$  as replacements for the constant volatility  $\sigma$  in the Black-Scholes formulas, e.g.,  $c(x) = c(x, \sigma, T)$  or  $F(x) = N(-d_2(x, \sigma, T))$ .

In an especially meaningful example, we have experimented with a class of smoothing algorithms used in "Implied volatility functions: Empirical tests," by B. Dumas, J. Fleming and R. E. Whaley, *J. Finance*, vol. 53, pp. 2059-2106, Dec. 1998. These authors fit an implied volatility curve  $\sigma(x)$ , for the purpose of setting up a "strawman" option price model for testing (and defeating) a theory regarding the role of volatility in option pricing. Their "strawman" option pricing model  $c(x)$  was obtained by putting the resulting smoothed curve back into the Black-Scholes call formula. It is a "strawman" *ad hoc* model, because no intuitive notion of stock volatility could possibly vary with strike price, which the stock never "sees." Nevertheless, their model performed admirably, surpassing in predictive power the highly regarded "implied tree" method. One possible explanation offered was that their model mimicked in a smooth way interpolation methods actually employed by practitioners in the options markets. (See the discussion of "Volatility matrices" in Hull, cited above.) Such an approach to option pricing seems ideal to us, because of its accuracy and because its underlying rationale represents a market view. Thus, we use the Dumas-Fleming-Whaley model for our own entirely different purpose, that of forecasting probability distributions. All that is necessary is to differentiate their call price model, which, conveniently for us, is a smooth function of strike price and other standard variables such as time, current stock price, and the risk-free rate of interest. The formula for the cdf  $F(x)$  is, as before, this derivative with 1 subtracted, or

$$F(x) = N(-d_2(x)) + \frac{\partial c}{\partial \sigma} \sigma'(x).$$

We can make this very explicit. We have

$$\frac{\partial c}{\partial \sigma} = x\sqrt{T}N'(-d_2(x))$$

where  $N'(z)$  denotes standard normal density, while  $\sigma'(x)$  may be computed by differentiating the Dumas-Fleming-Whaley fitted volatility curve. The latter has the form

$$\sigma(x, T) = a_0 + a_1x + a_2x^2 + a_3T + a_4T^2 + a_5xT.$$

The coefficients  $\{a_i\}$  are determined by regression. This kind of quadratic curve-fitting is easily implemented. Dumas-Fleming-Whaley impose a constraint to prevent their volatilities from going below 0 (or even below 0.01), and we have imposed further constraints on extrapolations (which we often carry out beyond the range of their tests), to ensure that the final cdf does not go below zero or above one. We have experimented with other variations on their basic approach, for example, using linear interpolation in the time domain, where we do not need to take derivatives. Our methods would, of course work, with any approach, possibly quite different, to volatility curve-fitting, though the general Dumas-Fleming-Whaley approach has many things going for it: accuracy, conformity to marketplace use of Black-Scholes, smoothness (differentiability, in particular), conformity to historical experience regarding the smile structure of volatility curves (especially important for extrapolation), and simplicity (which, beyond ease of implementation, helps avoid overfitting). These advantages are achieved in a probability context that was not considered in the paper where these volatility curves were introduced.

### 3 The multivariate case

The methods in the previous sections are capable of generating a display of raw or smoothed and extrapolated probability distributions for any optionable asset. Option prices are quoted on a large number of securities, as well as on certain indices, such as the S&P 500.

However, an investor would also like to know future probability distributions for:

- His or her entire portfolio;
- Mutual funds;
- A security without a quoted option;
- A security in a hypothetical scenario.

All of these questions involve considerations of several securities at once, and the probabilities of their simultaneous configuration of prices. This is clearly a consideration in the first two items above, but also enters in the third, where we would want to extract as much information as possible about the security without a quoted option price from those correlated with it that do have quoted options. Finally, in scenario analysis there are many questions that involve considering the probabilities of several security prices occurring at once, including changes in factors influencing the market that might be modeled by changes in a portfolio of those securities most affected. We will take all of these issues up in the remainder of this document, but for now we just try to give a basic introduction.

For a portfolio of securities, or a mutual fund, we are interested in a composite asset of the form

$$x = h_1x_1 + h_2x_2 + \cdots + h_nx_n,$$

where the  $x_i$  are all assets for which we individually know the cdf  $F(x_i)$  or the pdf  $f(x_i)$ . To give our method the most flexibility, we do not require that this knowledge come from any particular

procedure, though we favor the approach of the preceding two sections. However, even for some securities or indices with a quoted option, we might not feel there was sufficient option activity to justify a full fit of a volatility curve, and might take a cruder substitute, even a flat straight line based on an average of available implied volatilities. In addition, it is convenient to allow the possibility that a few assets we are monitoring might not have any quoted option at all; this is easily accommodated by, say, using a flat volatility curve with a historical value for volatility. For testing purposes and comparisons we might even want to consider a list of assets with all volatility curves given this way. In any case our methodology here is very general, and we only require that we know warping functions  $y_i(x_i)$  such that  $F(x_i) = N(y_i(x_i))$  for all  $i$ . If the asset has an active options market, then the warping functions may be determined by either first estimating  $F(x_i)$  directly from (finite differences of) options price data, as in subsection 2.2, or by using the approach discussed later in Section 2 of extrapolating and smoothing in the volatility domain. In the latter case we have an explicit form of the warping function  $y_i(x_i)$  in terms of a fitted volatility curve  $\bar{\sigma}_1(x_i, T)$  as  $y_i(x_i) = -d_2(x_i, \bar{\sigma}_1(x_i, T), T)$ , and this equation can also be used with any volatility curve with the assets above, that might have fewer or no traded options. In a later section we will discuss portfolios in the logarithm domain, possibly containing long and short positions. One can think of warping them to standard normal directly, subtracting the mean and dividing by the standard deviation. Alternately, to keep our notation uniform, one can invent an asset with price  $x_i$  such that  $-d_2(x_i)$  gives this warped value (using for  $\sigma_1(x)$  the observed historical volatility). But we wish to emphasize that the method we are describing works with ANY single-variable warping functions, even using a different one for each variable. The only further substantive ingredient is the plausibility of using JOINTLY normal distributions, which we now discuss.

The general problem is to find a multivariate probability distribution for the complete set of variables  $(x_1, \dots, x_n)$ , or equivalently for their logarithms. In simple financial models generalizing the Black-Scholes framework, the multivariate distribution of the logvariables is multivariate (i.e., jointly) normal; see Musiela and Rutkowski's book "Martingale methods in financial markets" (1999). This implies that all portfolios of these logvariables are jointly normal, and can also be used with other logvariables and portfolios of them to form a jointly normal distribution. Thus, if we wish, it is reasonable to use BARRA (or functionally equivalent) factors as single (log)variables in our model, using, say, individual normal distributions for them based on historical volatility. These factors may represent fundamentals of companies or even macroeconomic variables such as interest rates. We do not further discuss such factors, but refer to the book of Grinold and Kahn cited above, which also describes how to closely approximate them as portfolios of security returns. Our preference is to not use BARRA factors directly, but stay as much as possible in the world of optionable securities, and address questions involving BARRA factors in terms of approximating portfolios consisting mostly of optionable securities. (But for testing and comparisons, it is still useful to be able to include them directly, and we do have that capability.)

Now we certainly do not wish to use only the simple multidimensional Black-Scholes model, which would not directly allow the nonlognormal input from our single-variable distributions based on the options markets. At the same time, option prices on individual assets do not tell us anything about how assets interact, in particular, their correlations. Fortunately, correlations may be estimated from past (historical) data, and may be viewed as covariances for data that has been standardized (has standard deviation 1). Each multivariate normal distribution is determined by its mean and covariance matrix. Thus, a natural approach is to use the individual distributions to transform or "warp" the variables to standard normal, then impose a

multivariate normal structure based on the correlation matrix. This procedure is independent of the individual warping functions, which may be different for different individual variables, and in particular, can incorporate our market-based option distributions for individual variables representing securities with active options markets. A slightly different approach is to use correlations of the warped variables. This procedure is likely to be more accurate, but may involve more computational time.

We indicate some details. As before, it is notationally convenient to use  $v_i$  as a second notation for  $x_i$ , favoring the latter for fixed values and the former as a variable. Let  $C$  be the historical correlation matrix of the log variables  $(\ln v_1, \dots, \ln v_n)$ , whose entries are the cross-correlations

$$\rho_{ij} = \frac{E(\ln v_i \ln v_j) - E(\ln v_i)E(\ln v_j)}{\sqrt{E(\ln v_i)^2 E(\ln v_j)^2}}.$$

Then all diagonal terms  $\rho_{ii}$  are equal to 1, and  $C$  is a positive semi-definite covariance matrix, which we may here assume to be nonsingular (positive definite). If we use instead correlations of warped variables, we have simply

$$\rho_{ij} = E(y_i y_j).$$

Let us define  $F_C(y_1, \dots, y_n)$  as the cdf of a multivariate Gaussian random variable with mean zero and covariance matrix  $C$ . Thus,  $F_C(b_1, \dots, b_n)$  is the probability that each variable  $y_i$  is at most some value  $b_i$ . There are more elaborate versions, such as  $F_C(a_1, \dots, a_n; b_1, \dots, b_n)$ , giving the probability that each  $y_i$  satisfies  $a_i \leq y_i \leq b_i$ . In the single-variable case these latter functions are obtained from the simple cdf by a single subtraction, involving two terms, but the corresponding bivariate case involves four terms, and in  $n$  dimensions there would be  $2^n$  terms. However, each of these more elaborate cdf's can be directly computed as an integral, just like the simple cdf. Since the more elaborate cdf's are needed for Monte Carlo calculations, possibly in high dimensions, it is best to think of them as being computed directly.

We then define the multivariate cdf's

$$F(x_1, \dots, x_n) = F_C(y_1(x_1), \dots, y_n(x_n)),$$

and

$$F(a_1, \dots, a_n; b_1, \dots, b_n) = F_C(y_1(a_1), \dots, y_n(a_n); y_1(b_1), \dots, y_n(b_n))$$

where the  $y_i(x_i)$  are the known warping functions for the individual variables. We find it convenient, with some abuse of language, to speak of  $F(x_1, \dots, x_n)$  as "the cdf", even though we have all of the above functions in mind, and to use  $F(x_1, \dots, x_n)$  as a proxy for the whole distribution (which it does, theoretically, determine). This multivariate cdf then has the following properties:

- Since the marginals of  $F_C(z_1, \dots, z_n)$  are Gaussian with mean 0 and variance 1, the marginals of  $F(x_1, \dots, x_n)$  are equal to  $N(y_i(x_i)) = F(x_i)$ ; i.e., they are correct according to each single-variable model.
- If the logvariables  $(\ln v_1, \dots, \ln v_n)$  are actually jointly Gaussian, then the multivariate cdf  $F(x_1, \dots, x_n)$  is correct.

In summary, the true joint distribution is approximated by a jointly lognormal distribution using historical correlations, combined with warping functions on each variable such that the marginal distribution of each variable is correct according to a selected single-variable model

(for example, according to our single-variable model for optionable securities, or according to the lognormal model using historical volatility). The single variables may actually be portfolios, with a default distribution for the portfolio return being lognormal, based on historical volatility. This multivariate theory generalizes both our single-variable theory and standard multivariate (log)Gaussian models. It again allows for market input through option prices, to the extent that components have an active option market, but does not exclude nonoptionable securities, and also allows portfolios as single variables. In this way BARRA (or functionally equivalent) factors are also allowed because of their interpretation as portfolios of long and short positions.

## 4 Applications to portfolios

Given the multivariate cdf  $F(x_1, \dots, x_n) = F_C(y_1(x_1), \dots, y_n(x_n))$ , we can answer many typical questions. We first give an overview, and then take up some of the applications in more detail.

As one example, suppose that we want to find the cdf of a portfolio variable

$$x = h_1 x_1 + h_2 x_2 + \dots + h_n x_n,$$

where the  $h_i$  are arbitrary coefficients. A simple Monte Carlo method, probably not the fastest, is to draw random samples from the jointly Gaussian distribution with cdf  $F_C(y_1, \dots, y_n)$ , transform each  $y_i$  via the inverse mapping function  $x_i(y_i)$ , and then compute the resulting output sample

$$x = h_1 x_1(y_1) + \dots + h_n x_n(y_n).$$

After enough samples, we will have an approximation to the cdf of  $x$ . More precisely, the probability that  $a \leq x \leq b$  is, approximately, the average number of samples  $y_1, \dots, y_n$  with  $a \leq h_1 x_1(y_1) + \dots + h_n x_n(y_n) \leq b$ , and this approximation becomes exact in the limit for large sample sizes. This works for real portfolios, or for portfolios constructed from a number of assets and a residual variable, as might arise from a regression. Usually the regression is done in the log domain, which we discuss below. Note that the Monte Carlo method just described works perfectly well if the expression for  $x$  above is replaced by any function  $f(x_1, \dots, x_n)$  of the  $x_i$ , possibly quite nonlinear.

### 4.1 Log domain portfolios

In this subsection, we point out how our methods fit with another paradigm in common use in the financial community, and set up some further notation. It is common to work in the return domain, or equivalently, with logarithms; i.e.,

$$\ln x = \beta_1 \ln x_1 + \dots + \beta_n \ln x_n.$$

Ignoring any possible identification of these variables with those in the previous section, the same discussion and Monte Carlo method as above applies, if we regard  $x$  as a nonlinear portfolio  $x = f(x_1, \dots, x_n) = \exp(\beta_1 \ln x_1 + \dots + \beta_n \ln x_n)$ . If the sum  $B$  of the  $\beta_i$ 's is 1, such an  $x$  may be written  $x = \hat{h}_1 x_1 + \hat{h}_2 x_2 + \dots + \hat{h}_n x_n$  where  $\hat{h}_i = \beta_i x / x_i$ . Even if  $B$  is not 1, incremental changes ("returns")  $d \ln x$  computed from this equation for  $x$  are consistent with the above expression for  $\ln x$ . It is common in the financial community to think of  $\hat{h}_i$  as approximately a constant  $h_i$ , so

that for short periods, where the  $x_i$ 's do not change too much, this equation for  $x$  is comparable to the portfolio equation in the previous subsection.<sup>2</sup>

For an asset  $x$  not given explicitly in terms of the terms of the  $x_i$ , we obtain a similar expression via linear regression:

$$\ln x = \beta_0 \ln x_0 + \beta_1 \ln x_1 + \dots + \beta_n \ln x_n,$$

The  $\beta_i$  for  $i \neq 0$  are correlation coefficients chosen to minimize the variance of the residual in historical data (perhaps subject to constraints, such as  $\beta_i \geq 0$  and  $\sum_{i=1}^n \beta_i = 1$ ). For example,  $x$  might be an security without a quoted option, and the  $x_i$  for  $i \neq 0$  could be taken as assets for which we individually know the probability distributions, in addition to the required correlation coefficients for  $x$ . We have written the residual term as  $\beta_0 \ln x_0$  (usually thinking of  $\beta_0 = 1$  and the residual as normally distributed).<sup>3</sup> The mean of the latter could be nonzero, giving the regression "alpha"—a constant term making the mean of the regression correct. Alternatively, we could modify the equation to allow an explicit alpha, and keep the residual mean zero. Another minor variation might include the addition of a dummy variable with constant return, to adjust the value of  $x$  up or down. In particular, this gives another way of adjusting the residual mean to zero. This equation gives the previous one as a special case if we allow  $\beta_0 = 0$ .

#### 4.1.1 Fast fits of portfolios

One approach, which promises to be relatively fast computationally, is the following. As in the development of cdf-implied volatilities in Section 2, let us assume that each logvariable  $\ln x_i$  above is "Gaussian" with nonconstant variance  $\sigma_1(x_i)^2 T$ . In other words, the cdf is given by  $F(x_i) = N(-d_2(x_i, \sigma_1(x_i), T))$ . Our aim will be to give  $F(x)$  by a similar equation, using some kind of fitted curve  $\sigma_1(x)$ . We will assume that we have some class of volatility curves in mind, with a small number of parameters which must be determined.

If the variables  $\ln x_0, \dots, \ln x_n$  were truly jointly Gaussian, then  $\ln x$  would also be Gaussian. Its variance would be given by the formula

$$\text{Var}(\ln x) = \sum_{i,j} \beta_i \sigma_i \rho_{ij} \beta_j \sigma_j T,$$

where  $\rho_{ij}$  is the correlation between  $\ln x_i$  and  $\ln x_j$ , and  $\sigma_i^2 T = \text{Var}(\ln x_i)$ . We therefore define the estimate  $\hat{\sigma}_1(x)$  of  $\sigma_1(x)$  by the conditional expectation

$$\hat{\sigma}_1(x) = E(\beta_i \sigma_1(x_i) \rho_{ij} \beta_j \sigma_1(x_j) \mid \ln x = \sum_i \beta_i \ln x_i).$$

The calculation of the above conditional expectation may be done with Monte Carlo methods. In the language of nonlinear portfolios above, we would take the function  $f(x_1, \dots, x_n)$  to be 0 outside a thin multidimensional solid enclosing the hyperplane defined by  $\ln v = \sum_i \beta_i \ln v_i$ . Inside the solid we would take  $f(x_1, \dots, x_n)$  equal to the above expression for  $\text{Var}(\ln x)$ , divided by the probability of being in the solid (also a Monte Carlo calculation). In terms of samples, we just take the average of  $\text{Var}(\ln x)$  over all the samples that end up inside the thin solid. However,

<sup>2</sup>Thus  $\hat{h}_i \frac{dx_i}{x_i} = \beta_i \frac{dx_i}{x_i} = \beta_i d \ln x_i$ , so that for small changes  $dx_i$  the change  $dx$  from the first equation is approximately the same as would be obtained from the second. However, this relationship requires "rebalancing" to remain a good approximation for longer periods.

<sup>3</sup>For the residual term  $i = 0$ , we can use a constant variance, or impose some generic nonconstant structure based on observed behavior.

it is not necessary to compute all values of  $\hat{\sigma}_1(x)$ , but only enough to fit the parameters for the volatility curves we are using.

The estimated mean of  $\ln x$  would be  $\ln s^* - \frac{1}{2}\hat{\sigma}_1(x)^2$ , with  $s^*$  determined as before, or replaced with some risk-averse estimate, to obtain the risk-averse or "true" distribution. (It is common, incidentally, to use factor models such as these to estimate a risk-averse version of  $\ln s^* = \sum_i \beta_i \ln s_i^*$  from risk-averse values of  $s_i^*$ .)

Also, we mention here one useful variation: We may prefer not to view the residual term  $\beta_0 \ln x_0$  as part of the model, and instead write down a joint pdf only for  $\ln x_1, \dots, \ln x_n$ . In this case we can use the double expectation

$$\hat{\sigma}_1(x) = E(E(\sum_{i,j} \beta_i \sigma_1(x_i) \rho_{ij} \beta_j \sigma_1(x_j) \mid \ln x = \sum_i \beta_i \ln x_i)),$$

where the inner expectation is with respect to the variables  $x_1, x_2, \dots, x_n$ , and the outer expectation is with respect to the residual. We might take the standard deviation  $\sigma(x_0)$  of the residual (taking  $\beta_0 = 1$ ) as a constant, determined historically, or make an estimate based on some leverage model.

Now we can estimate the cdf  $F(x)$  by

$$\hat{F}(x) = N(-d_2(x, \hat{\sigma}_1(x), T))$$

as in the univariate case. To summarize, we use our multivariate model to determine parameters for a univariate model of the portfolio. After that is done, we can obtain probabilities for the portfolio without having to go back to the multivariate model, thus achieving a savings in time. We could take this one step further and think of randomly generating values of  $\sigma_1(x, T)$  independently of any Monte Carlo philosophy (but perhaps still throwing away values of  $x$  too far out-of-the-money), and then using the values obtained to do the regression required in the Dumas-Fleming-Whaley approach.

## 5 "What-if" questions

The multivariate distribution lends itself to the study of many questions regarding conditional probabilities. For example, suppose that we want to know the effect of the increase or decrease of some segment of the market on a portfolio, or the increase or decrease of some macro-economic factor. BARRA, following earlier ideas of Ross, has viewed such macro-economic factors as portfolios with both long and short positions. Similarly, BARRA considers market segments associated to price-to-earnings ratios and other fundamental parameters, as well as to industry groupings, as portfolios. (See the book of Grinold-Kahn cited above.) Thus, we are led simply to consider the effect of one portfolio on another.

For definiteness, let us suppose the first portfolio is  $x$ , where as above

$$\ln x = \beta_0 \ln x_0 + \beta_1 \ln x_1 + \dots + \beta_n \ln x_n,$$

and the second portfolio is  $y$ , where

$$\ln y = \gamma_0 \ln y_0 + \gamma_1 \ln x_1 + \dots + \gamma_n \ln x_n.$$

We take  $\beta_0 = \gamma_0 = 1$ , and view  $\ln x_0$  and  $\epsilon = \ln y_0$  as residuals with mean 0. The latter residual is not assumed to be a factor in our multivariate model. Consider the following typical "what-if"

question: Let  $A$  and  $B$  be given positive constants. If we know  $x \geq A$  at time  $T$ , what is the probability that  $y \geq B$  at time  $T$ ? We give two approaches to this problem, the first probably quicker, but possibly not as accurate, using a regression to avoid at least some Monte Carlo calculations.

### 5.1 "What-if": An approach involving part regression, part Monte Carlo

We have  $\ln y \geq \ln B$  iff  $\ln y - \epsilon \geq \ln B - \epsilon$ . All correlations  $\rho_{ij}$  between  $\ln x_i$  and  $\ln x_j$  are assumed known. We may also assume that we have historical values of volatilities  $\sigma_i = \sqrt{\text{Var}(\ln x_i)}$ . (Alternatively, we could estimate such values as expected values of implied volatilities, but it would not be difficult to maintain an inventory of historical values, and more in the spirit of this part of the calculation to do so.) Thus we can estimate the historical covariances between  $\ln x$  and  $\ln y - \epsilon$ :

$$\text{Cov}(\ln x, \ln y - \epsilon) \approx \sum_{i,j} \beta_i \sigma_i \rho_{ij} \gamma_j \sigma_j T,$$

as well as  $\sigma_{\ln x} = \sqrt{\text{Var}(\ln x)}$ ,  $\sigma_{\ln y - \epsilon} = \sqrt{\text{Var}(\ln y - \epsilon)}$  and the correlation

$$\rho(\epsilon) = \rho_{\ln x, \ln y - \epsilon} = \frac{\text{Cov}(\ln x, \ln y - \epsilon)}{\sigma_{\ln x} \sigma_{\ln y - \epsilon}}.$$

This gives a standard regression for the variable  $\ln y - \epsilon$  expressed in standard deviations from its mean, in terms of a similarly standardized expression for  $\ln x$ . Note that  $\epsilon$  has mean 0 by construction. Put  $d_2(s^*, x, \sigma) = \frac{\ln(s^*/x) - \sigma^2 T/2}{\sigma \sqrt{T}}$ . Thus  $-d_2(s_x^*, x, \sigma_{\ln x})$  measures standardized  $\ln x$  using historical volatility, and  $-d_2(x) = -d_2(s_x^*, x, \hat{\sigma}_1(x))$  measures "standardized" (warped)  $\ln x$  using the cdf-implied volatility curve  $\hat{\sigma}_1(x)$ , as discussed in the previous section. Here  $s_x^*$  denotes our best estimate for the value of  $x$  at time  $T$ .

Let  $\sigma_1(y, \epsilon)$  denote the volatility curve associated with  $\ln y - \epsilon$ , which may be estimated as in the previous section (or computed from estimates of  $\sigma_1(y)$  and the standard deviation of the residual, if we are willing to view the residual as uncorrelated with  $\ln y - \epsilon$ , as is guaranteed in unconstrained regression). Put  $d_2(y, \epsilon) = d_2(s_y^*, y e^{-\epsilon}, \hat{\sigma}_1(y, \epsilon))$ , so that  $-d_2(y, \epsilon)$  is a "standardized" measure of  $\ln y - \epsilon$ . Then the standard regression appropriate to our model is

$$-d_2(y, \epsilon) = \rho(\epsilon)(-d_2(x)).$$

There is a residual associated with this regression, which we have not written down. It is presumably normal, and its variance may be computed. For notational reasons we will just imagine it has been incorporated into the original  $\epsilon$ . As is apparent from the form of the expressions in the display, an alternative to the above regression is to do it with the warped correlation coefficients suggested in the previous section. If, in addition, it was appropriate to view the original portfolios as linear combinations of warped variables (our standard normal marginals), the regression above could be done without any recourse to Monte Carlo calculations. Similar remarks would apply if we used constant historical volatility functions throughout, though presumably the latter procedure would lose accuracy.

In any case, we can now answer our "what-if" question as a simple expectation in the univariate normal distribution of the (adjusted) residual  $\epsilon$ . Abbreviate  $d_2(s_x^*, A, \hat{\sigma}_1(A))$  to  $d_2(A)$  and



$d_2(s_y^*, B e^{-\epsilon}, \hat{\sigma}_1(B, \epsilon))$  to  $d_2(B, \epsilon)$ . Assume  $\rho(\epsilon) > 0$  (the natural case of a positive correlation). Then we have

$$\begin{aligned} \Pr\{y \geq B \mid x \geq A\} &= E(\Pr(-d_2(y, \epsilon) \geq -d_2(B, \epsilon) \mid -d_2(x) \geq d_2(A))) \\ &= E(\Pr(-d_2(x) \geq \rho(\epsilon)^{-1}(-d_2(B, \epsilon)) \mid -d_2(x) \geq -d_2(A))) \\ &= E(\min\{1, N(\rho(\epsilon)^{-1}(d_2(B, \epsilon))/N(d_2(A)))\}). \end{aligned}$$

The first equation follows just because  $-d_2(y, \epsilon)$  is monotonically increasing as a function of  $y$ ; that is, the condition that  $y \geq B$  is completely equivalent to the condition  $-d_2(y, \epsilon) \geq -d_2(B, \epsilon)$ . Similar remarks hold for the condition  $x \geq A$ , while the expression  $\Pr\{y \geq B \mid x \geq A\}$  just means the probability that the condition  $y \geq B$  holds when it is known that  $x \geq A$ . The second equation is then derived with the displayed expression above for  $-d_2(y, \epsilon)$ . (If  $\rho(\epsilon)$  is negative, the inequality involving its inverse reverses.) This inner expectation is then calculated in the normal distribution. For values of  $\epsilon$  for which  $-d_2(A)$  is as large as  $\rho(\epsilon)^{-1}(-d_2(B, \epsilon))$ , the expectation is a certainty, and yields the value 1. When  $-d_2(A)$  is smaller than  $\rho(\epsilon)^{-1}(-d_2(B, \epsilon))$ , its cumulative normal distribution value  $N(-d_2(A))$  is smaller than  $N(\rho(\epsilon)^{-1}(-d_2(B, \epsilon)))$ , and the probability  $1 - N(-d_2(A)) = N(d_2(A))$  that the standard normal variable  $z = -d_2(x)$  is at least  $-d_2(A)$  is smaller than the corresponding probability  $1 - N(\rho(\epsilon)^{-1}(-d_2(B, \epsilon))) = N(\rho(\epsilon)^{-1}(d_2(B, \epsilon)))$  that  $z$  be at least  $\rho(\epsilon)^{-1}(-d_2(B, \epsilon))$ . The ratio  $N(\rho(\epsilon)^{-1}(d_2(B, \epsilon))/N(d_2(A)))$ , which is the desired inner expectation, is thus smaller than 1, as is appropriate for a probability, conditional or not. If  $\rho(\epsilon)$  is negative, similar reasoning leads instead to the expression  $E(\max\{0, (N(d_2(A)) - N(\rho(\epsilon)^{-1}(d_2(B, \epsilon))))/N(d_2(A))\})$  for the desired conditional probability. Although the final answer in either case is an expectation (over  $\epsilon$ ), it is essentially an integral that could be computed quickly with power series. (A very simple and accurate power-series expansion of  $N(z)$  is given on p. 252 of the book by Hull cited above.) Using that, one could determine by iterative methods what value of  $\epsilon$  makes, say, the ratio  $N(\rho(\epsilon)^{-1}(d_2(B, \epsilon))/N(d_2(A)))$  equal to 1, and then integrate the ratio against the standard normal pdf from  $-\infty$  to the determined value of  $\epsilon$ , in the  $\rho(\epsilon) > 0$  case. Similar remarks apply if  $\rho(\epsilon) < 0$ . (Note that, if  $\rho(\epsilon) = 0$ , the variables  $\ln x$  and  $\ln y$  are uncorrelated, and the conditional probability  $\Pr\{y \geq B \mid x \geq A\}$  is the same as the unconditional probability  $\Pr\{y \geq B\}$ .)

All of the latter calculations can be done very fast. Of course, we have already used some Monte Carlo calculations to get this far, unless we are in the simplified context of constant volatility functions.

## 5.2 "What-if": The full Monte

It is easy to say how we would compute an answer to the same "what-if" question, using our full joint probability distribution. We simply write

$$\Pr\{y \geq B \mid x \geq A\} = E(\Pr(-d_2(y, \epsilon) \geq -d_2(B, \epsilon) \mid -d_2(x) \geq -d_2(A)))$$

and interpret  $\ln x$  in  $-d_2(x)$ , and  $\ln y - \epsilon$  in  $-d_2(y, \epsilon)$  in terms of their expansions in  $\ln x_0, \ln x_1, \dots, \ln x_n$ . To compute, say the inner expectation by a Monte Carlo calculation, we would generate a large number of random samples of multivariate standard normal vectors  $z$  with covariance matrix  $C$ . We then take the average, over the samples  $z$  which happen to satisfy  $z \geq -d_2(A)$ , of the function which is 1 when  $-d_2(y, \epsilon) \geq -d_2(B, \epsilon)$  and 0 otherwise. We have not experimented to see whether this method yields better answers than the regression procedure above. Nevertheless, it illustrates how we could approach more sophisticated "what-if" questions

that could not be easily treated by regressions. For example, suppose we believe that factor  $w$  will remain in a range  $C \leq w \leq D$ , and ask the same question about  $y$ , subject to the same condition on  $x$ . This is hard to formulate in terms of regression, and is simply not possible in terms of single-factor regression. However, it is easy to answer with the full distribution:

$$\Pr\{y \geq B \mid x \geq A, C \leq w \leq D\} = \\ E(\Pr(-d_2(y, \epsilon) \geq -d_2(B, \epsilon) \mid -d_2(x) \geq d_2(A)), -d_2(C) \leq -d_2(w) \leq -d_2(D)).$$

Finally, we may not want to work in the log domain, which, if we started with a fixed portfolio  $x = h_1 x_1 + h_2 x_2 + \dots + h_n x_n$  would force us into an approximation, as noted. But, working with the full distribution, we can phrase a condition  $x \geq A$  as  $h_1 x_1(y_1) + \dots + h_n x_n(y_n) \geq A$ , in the language of the first section where the vector of  $y$ 's plays the role of our vector  $z$  here. Monte Carlo calculations can now proceed as before, using log domain expressions or not for the other conditions.

## 6 "You gotta believe" questions

In the previous section we were focusing on an investor thinking about the value of his or her portfolio  $y$  in response to the change in a factor  $x$ . Conversely, an investor might want to know what the investment world looks like if a given stock or index  $y$  goes to a certain level  $B$  at time  $T$ . What is the expected value  $A$  at time  $T$  of another portfolio  $x$ , or simply of one of the factors  $x_i$ ? Our main plan is, upon input by the user that  $y$  is going to level  $B$ , to list several assets  $x_i$  or factors/indices  $x$  most highly correlated with  $y$  and their expected values with  $y$  at  $B$ .

It would also be possible to display a confidence interval for each selected asset or factor, and have other information about its new projected probability distribution readily available. We could also offer comparisons with the old projected probability distribution of  $x$ , where no assumptions on  $y$  is made. Finally, in some cases, where it was possible to explain much of the variance of  $y$  with just a few  $x_i$  (appearing in the regression of  $y$ ), we could list percentage increases/decreases of a portfolio of these  $x_i$  required to make  $B$  the expected value of  $y$ , based solely on its dependence on this portfolio. (For example, the coefficients in the portfolio could come from the regression of  $y$  with respect to all the  $x_i$ , or some new regression might be done, perhaps allowing user-defined constraints). It should be mentioned that medians or modes are alternatives to expected values (means) here and above; in any case users will need to be educated about the fact that the median and mode differ systematically from the mean in near lognormal distributions.

The main problem might be viewed as understanding the probability distribution of  $x$ , given that  $y \geq B$  at a given time  $T$ , with  $x$  and  $y$  as in the previous section. This can be approached by the methods of the previous sections, by reversing the roles of the variables.

There is, however, a simpler question that can be treated in an especially quick way. Consider the problem of determining the mean of  $x$  conditioned on the equality  $y = B$  at time  $T$ . The idea is to use simple regression methods, but interpret answers as measured in terms of our variable volatilities. In our previous notation, we have a regression

$$-d_2(x) = \rho \cdot (-d_2(y, \epsilon)) + v$$

where  $\rho$  (which we called  $\rho(\epsilon)$  in the previous section) is the historically determined correlation between  $\ln x$  and the random variable  $\ln y - \epsilon$ . Note that the roles of dependent and independent

variable are reversed. There is also a residual  $v$ , which has mean 0 here, and plays a role (gets averaged away). Thus, the desired conditional expected value  $A$  of  $x$  is obtained from

$$-d_2(A) = E(-d_2(x) | y = B) = \rho \cdot E(-d_2(s_y^*, B e^{-\varepsilon}, \hat{\sigma}_1(y, \varepsilon))) = \rho \cdot (-d_2(s_y^*, B, \hat{\sigma}_1(y, \varepsilon))).$$

Recall that  $\hat{\sigma}_1(y, \varepsilon)$  is an estimate, obtained by Monte Carlo methods, of the implied volatility  $\sigma_1$  associated to the random variable  $\ln y - \varepsilon$ . For faster but less accurate calculations it can be estimated historically as  $\sum_{i,j} \beta_i \sigma_i \rho_{ij} \beta_j \sigma_j$  with each of the  $\sigma$ 's,  $\beta$ 's, and  $\rho$ 's here given historically. (See the previous section for notation.) Similarly, for fast calculations,  $-d_2(\bar{x})$  could use historical volatility, though we expect it to be given more accurately, or rather, more accurately according to the market view, as  $-d_2(x) = -d_2(s_x^*, x, \hat{\sigma}_1(x))$ , using the implied volatility function estimate  $\hat{\sigma}_1(x)$ . If  $x = x_i$  is a single asset or index in our model, then  $\hat{\sigma}_1(x) = \hat{\sigma}_1(x_i)$  does not require a Monte Carlo estimate, but is presumably already available.

To summarize, the conditional expected values required to answer "you gotta believe" questions are easily obtained by regression methods. The accuracy of such answers is enhanced, or at least shaped more to reflect market input, when all logvariables are measured in "standard deviations," interpreted as our variable volatilities.

## 7 Portfolios containing option securities

We conclude this document by briefly pointing out that our methods, when using full Monte Carlo calculations, easily apply to portfolios containing option securities. The well-known idea is to think of an option as as a kind of nonlinear portfolio— a quadratic one, to be more precise. Thus, an option on a single underlying security with underlying price  $x_1$  has a price approximately  $x = c + \Delta(x_1 - s_1) + (1/2)\Gamma(x_1 - s_1)^2$  for  $x_1$  near  $s_1$ , where the option was evaluated to a known value  $c$ . Here  $\Delta$  and  $\Gamma$  are well-known parameters in the options markets, giving the first and second derivatives of the option price at  $s_1$  with respect to the underlying security price  $x_1$ . Perhaps the most characteristic feature of options is that they have nonzero  $\Gamma$ — their proportion of increase or decrease with respect to the underlying security price changes as the security price changes. Explicit formulas in terms of other standard parameters are available, say, in the Black-Scholes theory for both  $\Delta$  and  $\Gamma$  (see the Hull book cited above). Such formulas could be obtained by differentiation directly in other theories or when using empirically-fitted curves. In any case, once we have such an explicit approximation to  $x$ , its probability distribution is easily given by the Monte Carlo methods of Subsection 3.1 above. The same method applies as well to portfolios containing several options and other securities.

**CLAIMS**

- 1     1.     A method comprising:  
2             receiving data representing current prices of options on a given  
3     asset,  
4             deriving from said data an estimate of a corresponding implied  
5     probability distribution of the price of said asset at a future time, and  
6             making information about said probability distribution available  
7     within a time frame that is useful to investors.
- 1     2.     The method of claim 1 in which the data represent a finite number  
2     of prices of options at spaced-apart strike prices of the asset, and also  
3     including  
4             calculating a set of first differences of said finite number of prices  
5     to form an estimate of the cumulative probability distribution of the price  
6     of said asset at a future time.
- 1     3.     The method of claim 2 also including  
2             calculating a set of second differences of the finite number of  
3     strike prices from the set of first differences to form an estimate of the  
4     probability distribution function of the price of said asset at a future time.
- 1  
1     4.     A method comprising:  
2             receiving data representing current prices of options on a given  
3     asset,

4 deriving from said data an estimate of a corresponding implied  
5 probability distribution of the price of said asset at a future time, and  
6 providing a real-time data feed containing information based on  
7 said probability distribution.

---

1 5. A method comprising:

2 providing a graphical user interface for viewing pages containing  
3 financial information related to an asset; and

4 when a user indicates an asset of interest, displaying probability  
5 information related to the price of the asset at a future time.

1 6. A method comprising:

2 enabling a user to identify an asset of interest, the asset being one  
3 for which data representing current prices of options on the asset are  
4 available,

5 deriving from said data an estimate of a corresponding implied  
6 probability distribution of the price of said asset at a future time, and

7 providing a display of a probability distribution of prices of the  
8 asset at future times.

1 7. A method comprising:

2 enabling a user to indicate a future time and to identify an asset of  
3 interest, the asset being one for which data representing current prices of  
4 options on the asset are available, and

5 displaying to the user a distribution of the probability that the asset  
6 will reach prices within a range of prices at the future time.

1 8. A method comprising:

2 receiving data representing current prices of options on a given  
3 asset, the options being associated with spaced-apart strike prices of the  
4 asset at a future time,

---

5 the data including shifted current prices of options resulting from a  
6 shifted underlying price of the asset, the amount by which the asset price  
7 has shifted being different from the amount by which the strike prices are  
8 spaced apart, and

9 deriving from said data an estimate of a quantized implied  
10 probability distribution of the price of said asset at a future time, the  
11 elements of the quantized probability distribution being more finely  
12 spaced than for a probability distribution derived without the shifted  
13 current price data.

1 9. A method comprising

2 receiving data representing current prices of options on a given  
3 asset, the options being associated with spaced-apart strike prices of the  
4 asset at a future time,

5 deriving from said data an estimate of an implied probability  
6 distribution of the price of said asset at a future time, the mathematical  
7 derivation including a smoothing operation, and

8 making information about said probability distribution available  
9 within a time frame that is useful to investors.

1 10. The method of claim 9 in which the smoothing operation is  
2 performed in a volatility domain.

1 11. The method of claim 9 in which the smoothing operation is  
2 performed in the domain of the option prices or in the domain of the  
3 probability distribution information.

1 12. A method comprising:

---

2 receiving data representing current prices of options on a given  
3 asset, the options having strike prices at future dates,

4 deriving a volatility for each of the future dates in accordance with  
5 a predetermined option pricing formula that links option prices with strike  
6 prices of the asset;

7 generating a smoothed and extrapolated volatility function;

8 and using the volatility information to generate information within  
9 a time-frame that is useful for investors.

1 13. The method of claim 12 in which the volatility function is  
2 extrapolated to a wider range of dates than the future dates.

1 14. The method of claim 12 in which the volatility function is  
2 extrapolated to strike prices other than the strike prices of the options.

1 15. The method of claim 9 also including

2 generating a smoothed volatility function using only data that are  
3 reliable under a predetermined measure of reliability.

1 16. The method of claim 9, further comprising:

2 generating an implied volatility function formula having a  
3 quadratic form with two variables representing a strike price and an  
4 expiration date;

5 wherein coefficients of the implied volatility function formula are  
6 determined by applying regression analysis to approximately fit the  
7 implied volatility function formula to each of the implied volatilities.

8 17. A method comprising:

---

9 receiving data representing current prices of options on assets  
10 belonging to a portfolio,

11 deriving from said data an estimate of an implied multivariate  
12 distribution of the price of a quantity at a future time that depends on the  
13 assets belonging to the portfolio, and

14 making information about said probability distribution available  
15 within a time frame that is useful to investors.

1 18. A method comprising:

2 receiving data representing values of a set of factors that influence  
3 a composite value,

4 deriving from said data an estimate of an implied multivariate  
5 distribution of the price of a quantity at a future time that depends on  
6 assets belonging to a portfolio, and

7 making information about said probability distribution available  
8 within a time frame that is useful to investors.

1 19. The method of claim 18 in which the mathematical derivation  
2 includes generating a multivariate probability distribution function based  
3 on correlations among the factors.

1 20. A graphical user interface comprising:



2 a user interface element adapted to enable a user to indicate a  
3 future time;

4 a user interface element adapted to show a current price of an  
5 asset; and

---

6 a user interface element adapted to show the probability  
7 distribution of the price of the asset at the future time.

1 21. A method comprising:

2 continually generating current data that contains probability  
3 distributions of prices of assets at future times,

4 continually feeding the current data to a recipient electronically,  
5 and

6 the recipient using the fed data for services provided to users.

1 22. A method comprising:

2 receiving data representing current prices of options on assets  
3 belonging to a portfolio,

4 receiving data representing current prices of market transactions  
5 associated with a second portfolio of assets, and

6 providing information electronically on the probability that the  
7 second portfolio of assets will reach a first value given the condition that  
8 the first portfolio of assets reaches a specified price at a future time.

1 23. A method comprising:

2 receiving data representative of actual market transactions

3 associated with a first portfolio of assets;

4 receiving data representative of actual market transactions  
5 associated with a second portfolio of assets;

6 providing information on the expectation value of the price of first  
7 portfolio of assets given the condition that the second portfolio of assets  
8 reaches a first specified price at a specified future time through a network.

1

1 24. A method comprising

2 evaluating an event defined by a first multivariate expression that  
3 represents a combination of macroeconomic variables at a time T, and

4 estimating the probability that a second multivariate expression  
5 that represents a combination of values of assets of a portfolio will have a  
6 value greater than a constant B at time T if the value of the first  
7 multivariate expression is greater than a constant A.

1 25. The method of claim 24 in which the probability is estimated using  
2 Monte Carlo techniques.

1 26. A method comprising

2 defining a regression expression that relates the value of one  
3 variable representing a combination of macroeconomic variables at time T  
4 to a second variable at time T that represents a combination of assets of a  
5 portfolio, and

6 estimating the probability that the second variable will have a  
7 value greater than a constant B at time T if the value of the first variable is  
8 greater than a constant A at time T, based on the ratio of the probability of

9 x being greater than A under the regression expression and the probability  
10 of x being greater than A.

1 27. A method comprising

2 defining a current value of an option as a quadratic expression that  
3 depends on the difference between the current price of the option and the  
4 current price of the underlying security, and

5 using Monte Carlo techniques to estimate a probability distribution  
6 of the value at a future time T of a portfolio that includes the option.

1 28. A method comprising

2 displaying to a user a circular visualization element having sectors  
3 arranged around a center of the element, the sectors respectively  
4 corresponding to different groups of assets,

5 in each of the sectors, displaying an array of visual elements representative  
6 of respective assets belonging to the group to which the sector  
7 corresponds, the visual elements being arrayed with respect to distance  
8 from the center in accordance with magnitudes of performance of the  
9 assets during a recent period.

1 29. The method of claim 28 in which the visual elements comprise  
2 displayed dots, one for each of the assets.

1 30. The method of claim 28 in which the visual elements exhibit  
2 visible characteristics that correspond to categories of the assets within the  
3 group.

1 31. The method of claim 30 in which the categories of the assets  
2 within the group correspond to different capitalizations.

1 32. The method of claim 29 in which dots are arranged along a radius  
2 of the sector to which they belong.

1 33. The method of claim 32 in which dots that would otherwise lie on  
2 the radius at a given distance from the center are displayed at different  
3 angular positions near to the radius.

1 34. The method of claim 28 in which the sectors have angular extents  
2 that represent the fractions of the total number of asset items represented  
3 by the respective sectors.

1 35. The method of claim 28 in which the circular visualization element  
2 is subdivided into rings having respectively different distances from the  
3 center.

1 36. The method of claim 35 in which the rings are displayed in  
2 different colors.

1 37. The method of claim 28 in which the magnitudes of performance  
2 of the assets are measured in percentage price change.

1 38. The method of claim 28 in which the recent period comprises a  
2 trading day on an asset market.

1 39. The method of claim 28 in which the assets comprise securities  
2 issued by corporations.

1 40. A displayed visualization element that  
2 is circular,

3 has sectors arranged around a center of the element, the sectors  
4 respectively corresponding to different groups of securities issued by  
5 corporations,

6 in each sector, has an array of dots representing respective securities  
7 belonging to the group to which the sector corresponds, each of the dots  
8 lying on or near a radius of the sector and each having a distance from the  
9 center along the radius that corresponds to the percentage change in the  
10 price of the represented security during a trading day, and

11 has differently colored rings at respectively different distances from the  
12 center.

1 41. A method comprising

2 displaying to a user a visualization element that indicates the odds  
3 of a performance measure of an asset being within specified ranges of  
4 identified values of the performance measure at a succession of times in  
5 the future.

1 42. The method of claim 41 in which the performance measure  
2 comprises a price of the asset.

1 43. The method of claim 41 in which the performance measure  
2 comprises a return percentage.

1 44. The method of claim 41 in which the performance measure  
2 comprises a tax-adjusted return percentage.

1 45. The method of claim 41 in which the visualization element include  
2 stripes superimposed on a graph of the performance measure over time,  
3 each of the stripes representing one of the specified ranges.

1 46. The method of claim 45 in which each of the stripes begins at a  
2 current time and becomes broader as it extends to future times.

1 47. The method of claim 41 also including

2 displaying a graphical device that shows actual historical values of  
3 the performance measure.

1 48. The method of claim 47 in which the graphical device that shows  
2 actual historical values is a line graph one end of which joins the  
3 visualization element at a point which represents a current date.

1 49. The method of claim 41 in which the visualization element  
2 includes two portions, one of the portions representing the odds prior to a  
3 specified date based on one assumption, the other of the portions  
4 representing the odds after the specified date based on another assumption.

1 50. The method of claim 49 in which the specified date is a date on  
2 which tax effects change from the one assumption to the other assumption.

1 51. A method comprising  
2 displaying to a user a visualization element having graphical  
3 indicators of the relative performance of a selected asset compared with  
4 the performance of groups of assets in each of a succession of time  
5 periods, each of the groups comprising assets representing a common  
6 style.

1 52. The method of claim 51 in which the style comprises a class of  
2 investment objectives.

1 53. The method of claim 51 in which the relative performance is  
2 determined using an asset class factor model.

1

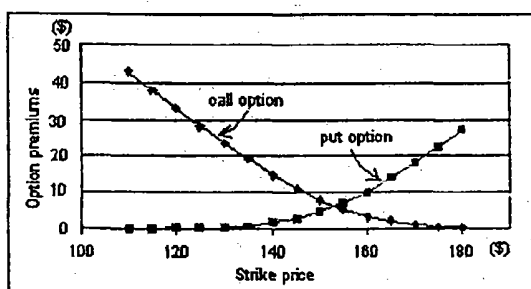


FIGURE 1

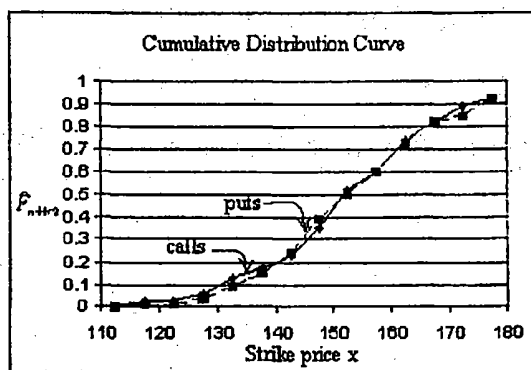


FIGURE 2

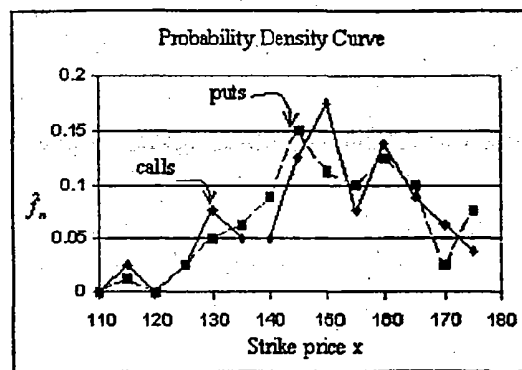


FIGURE 3

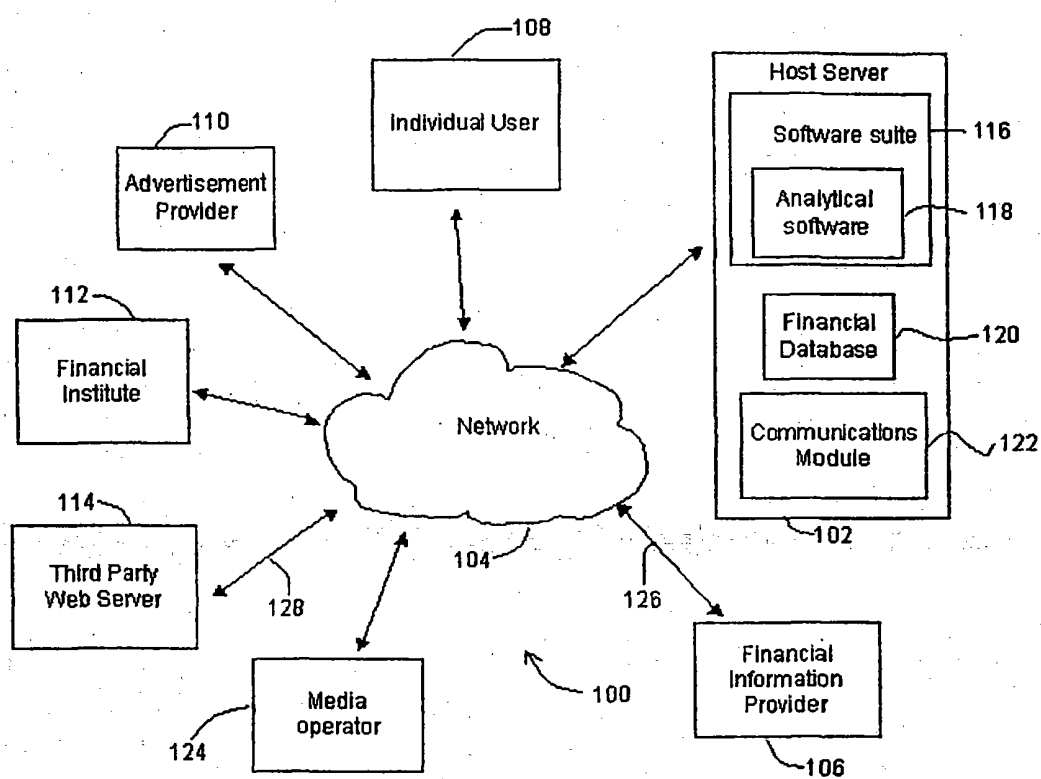


FIGURE A



<http://www.thinkinginvestments.com/rockets/newdemo/morningline.htm> - Microsoft Internet Explorer

File Edit View Favorites Tools Help

THINKING INVESTMENTS

Home My Financial Life Advice University Thinking Trader Portfolios Research

Membership | [About Us](#) | [Products](#) | [? Help](#) [Log In](#) [Log Out](#)

**Click here to learn more.**

Requests for more information, including changes to expenses. Read carefully before investing. Past performance is no guarantee of future results. Legg Mason Fund Watch, Inc.

**LEGG MASON FUNDS**

**ClearFuture<sup>SM</sup>**

We can help you reach your retirement goals with tailored 401(k) advice. Click here!

### The Morning Line

Find Symbol **720**

Historical Results **712**

Create / Edit Portfolio **716**

Select Portfolio/Watch List **710**

Select Time Horizon **3 Months**

Select View **Percents**

Today's Morning Line ☒

Symbols	Name	Alert	Price	Price Reference	Up%	Down%
SPX	S&P 500		1458.68	Current	50	50
CMGI	CMGI		46.50	46.50	30%	25%
MSFT	MICROSOFT		76.54	76.54	25%	15%
MOT	MOTOROLA		30.50	30.50	20%	10%
QCOM	QUALCOMM		60.50	60.50		

Visit our Sponsors

\* TR = Trace chance = Less than a 1% chance of this occurring

Alert Codes: **TR** = News Today; **H/L** = 52 week high/low; **E** = Earnings Report; **P** = Price Change

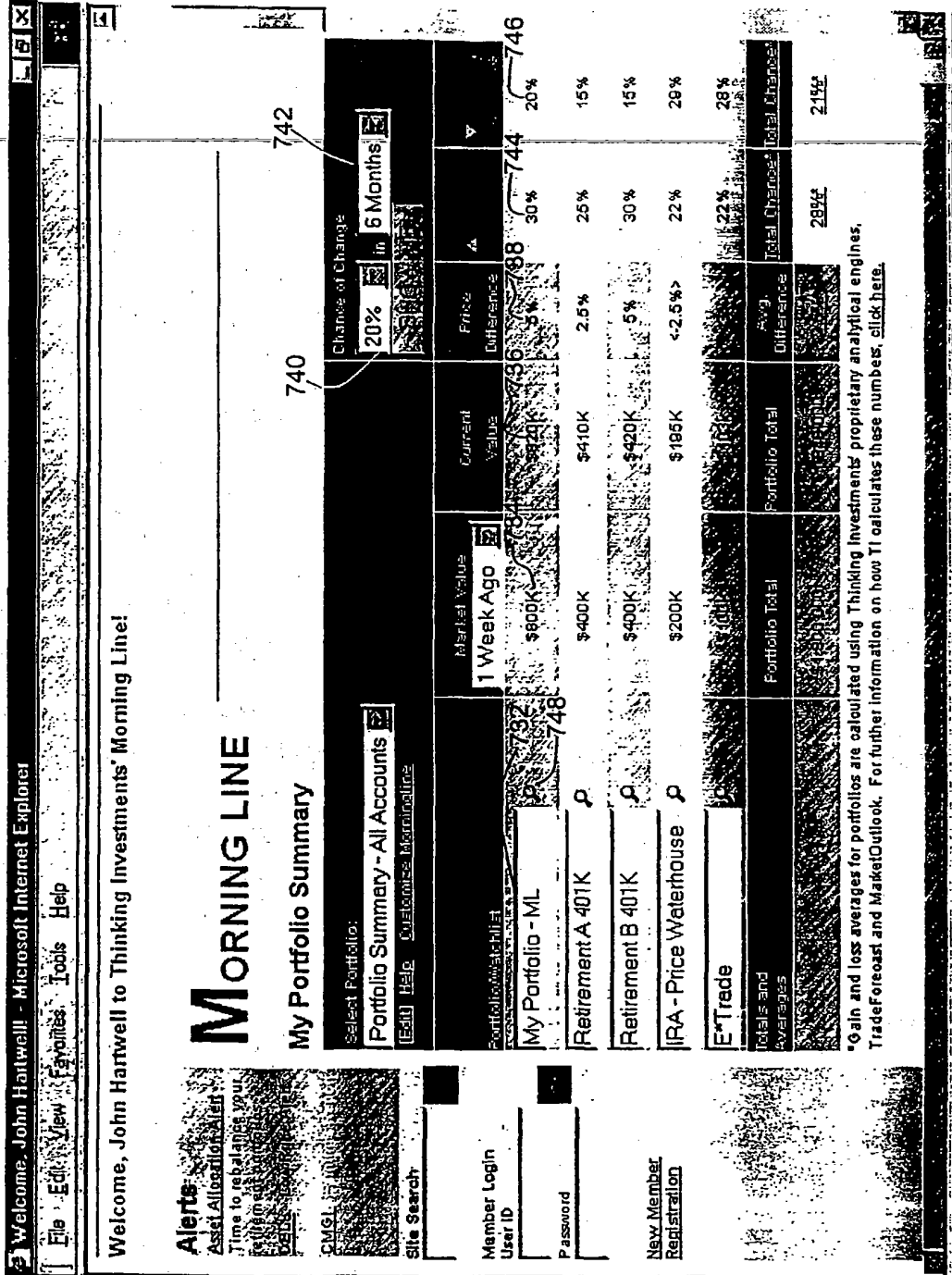
726

728

730

700

FIGURE 5



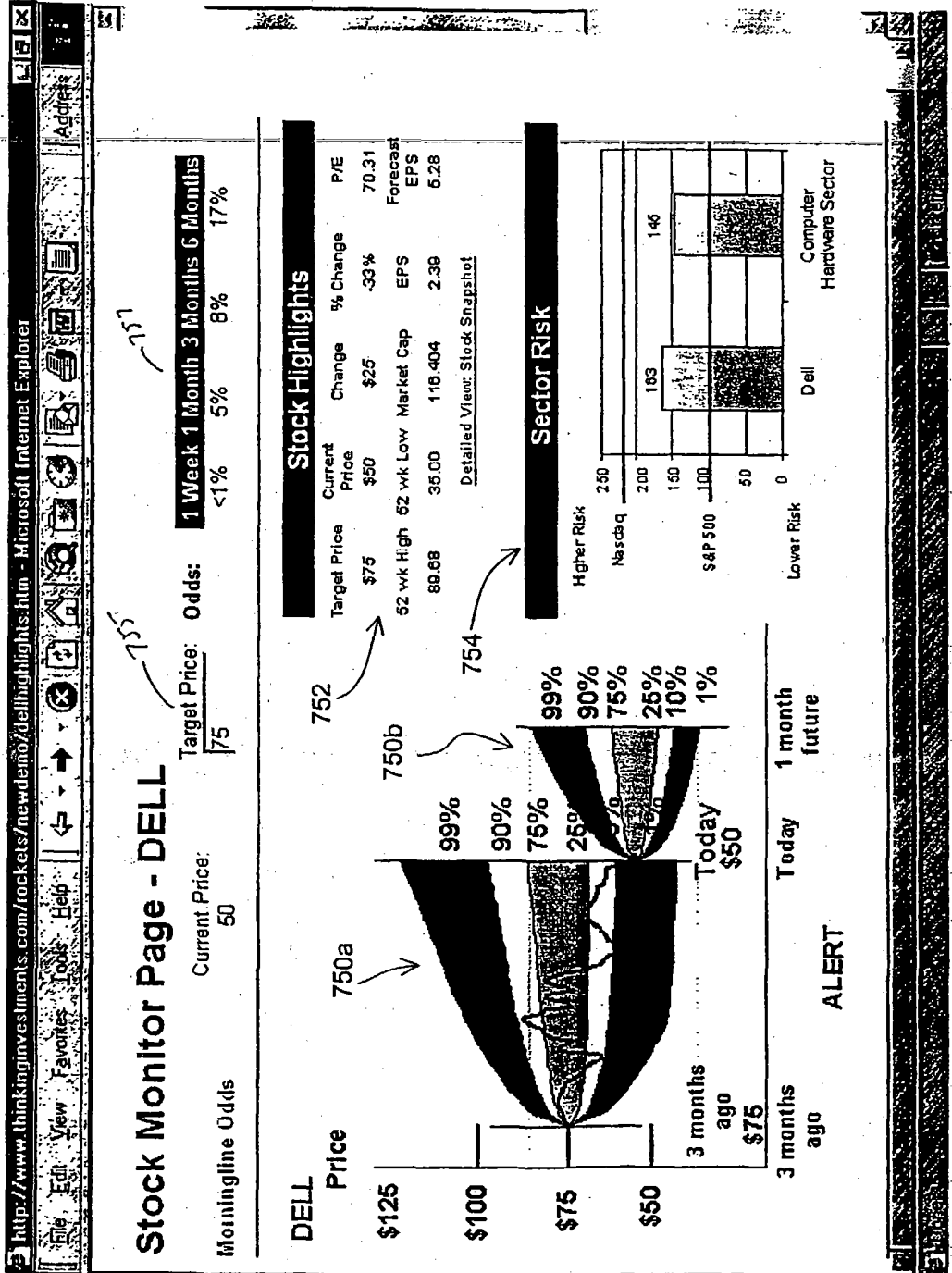


FIGURE 7

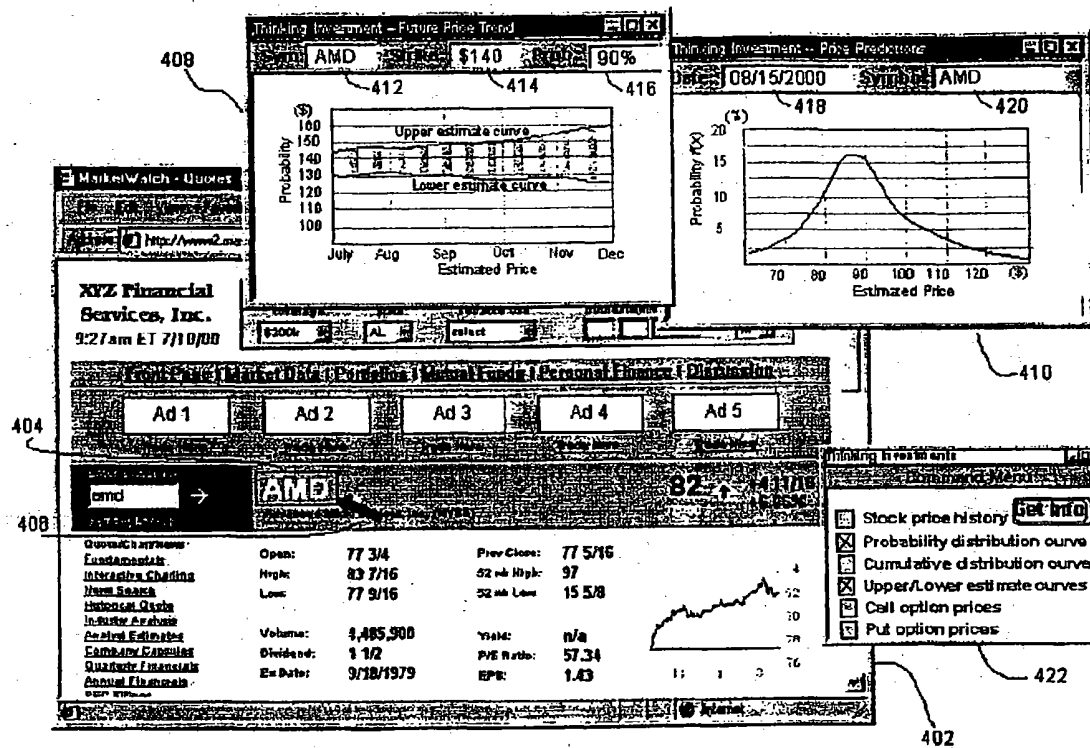


FIGURE 8

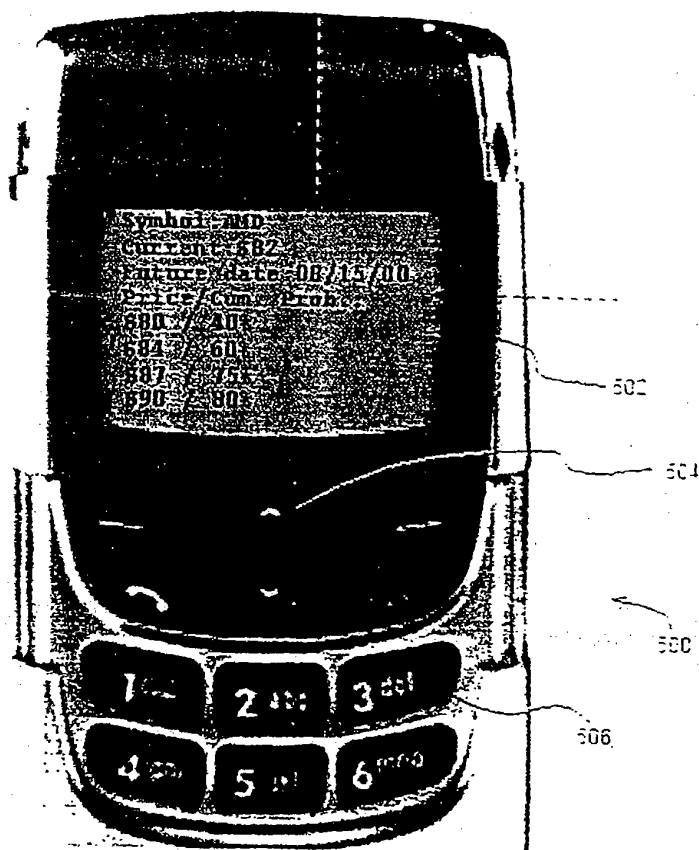


FIGURE 9

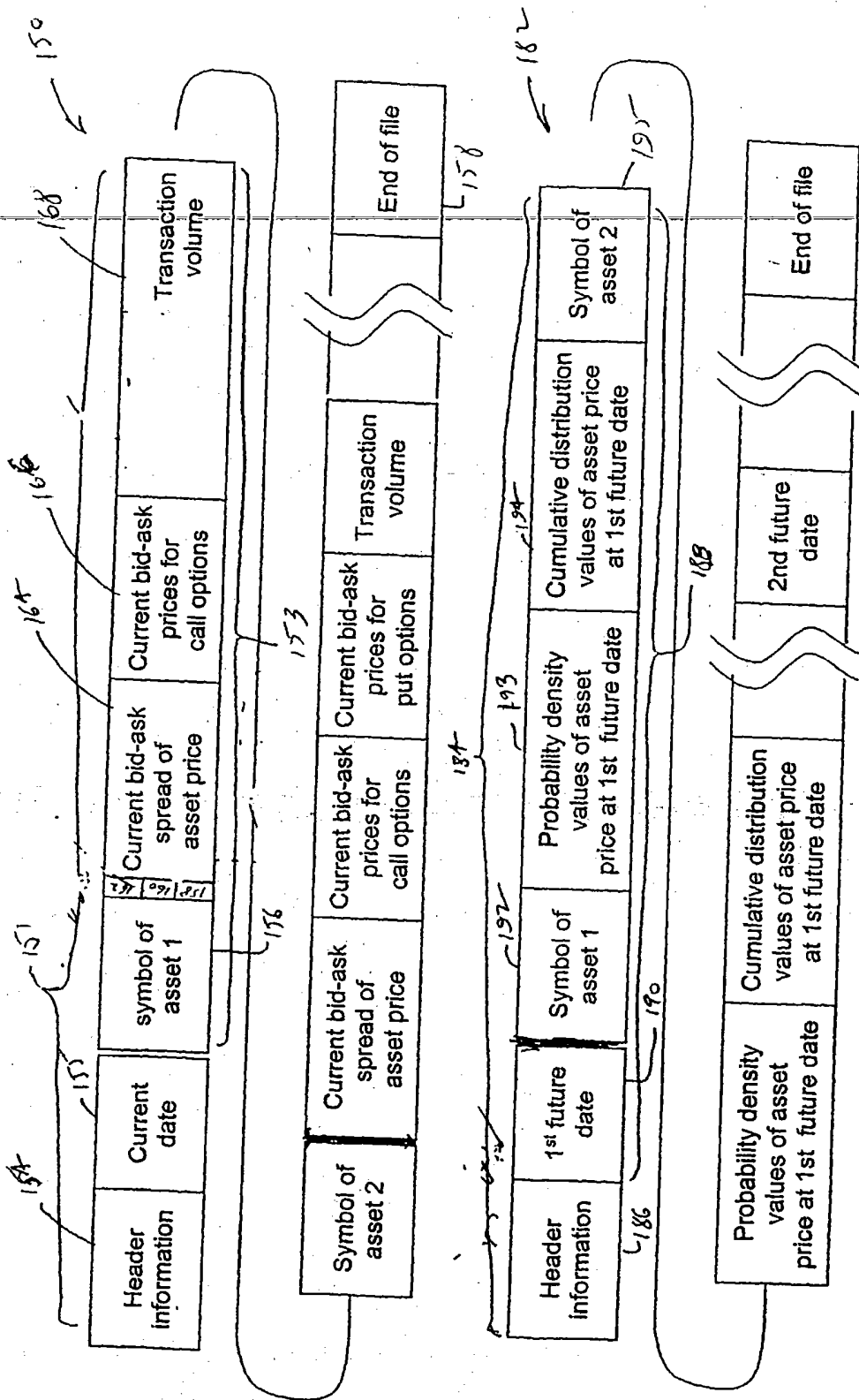


FIGURE 10

## FIGURES

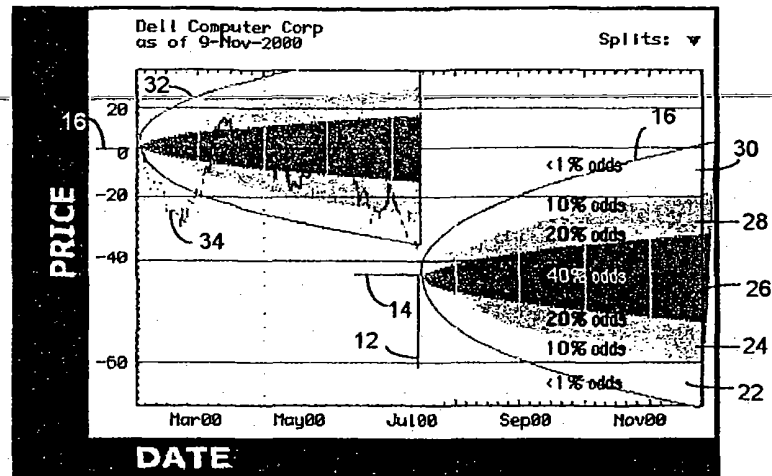


FIGURE 11

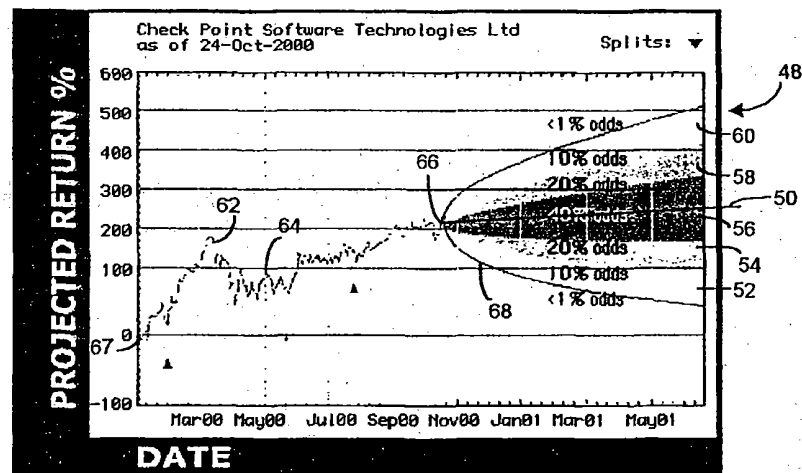


FIGURE 12

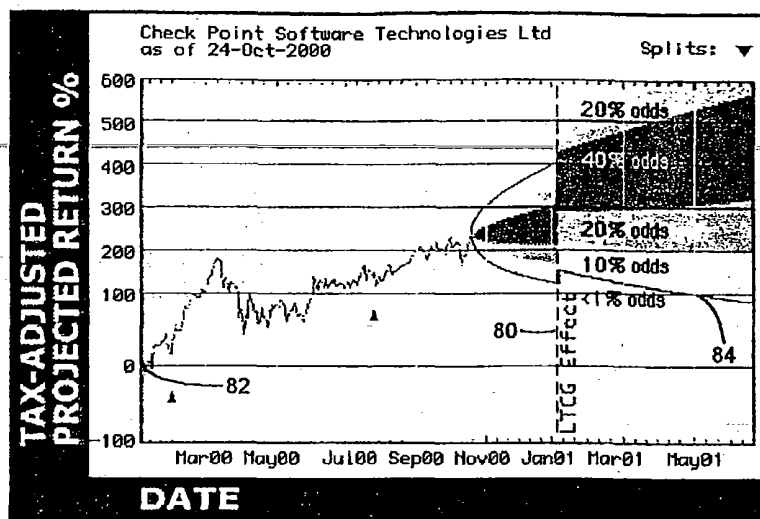


FIGURE 13



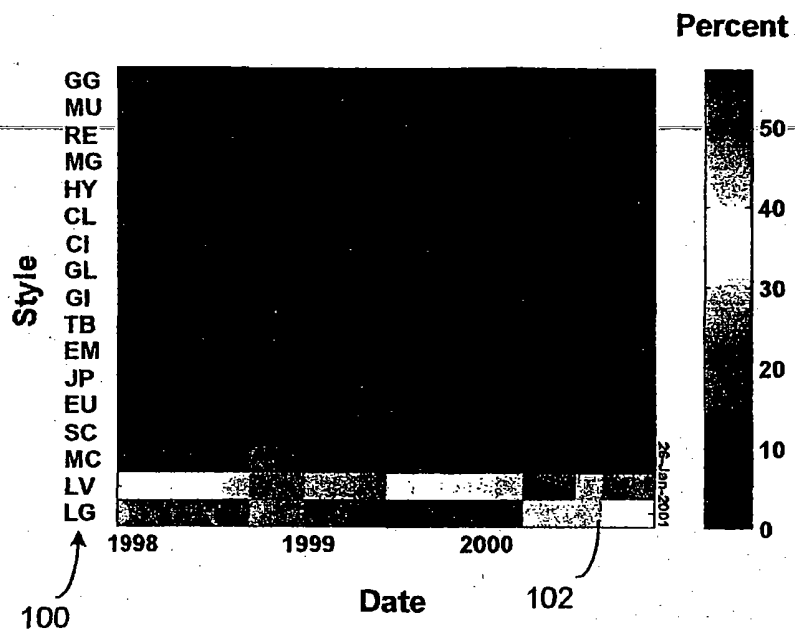


FIGURE 14

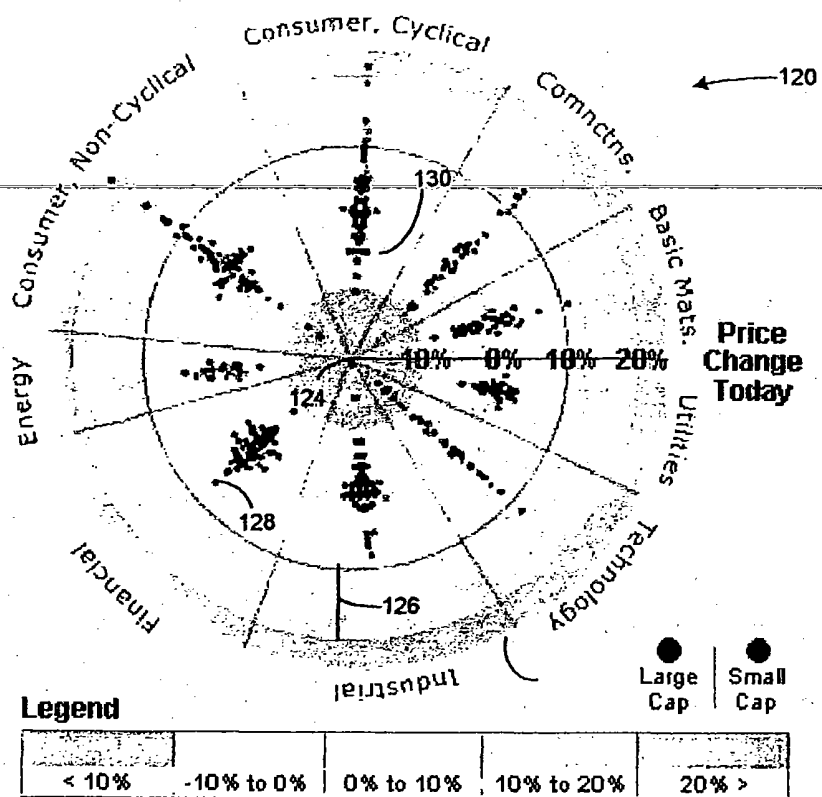


FIGURE 15

## FIGURES

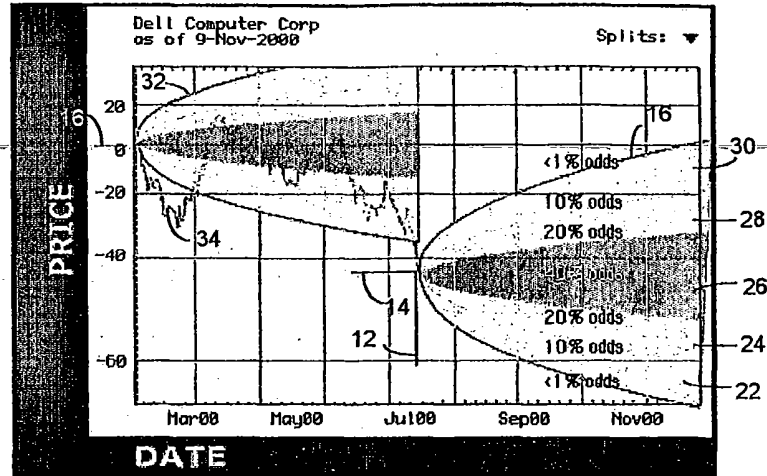


FIGURE 11

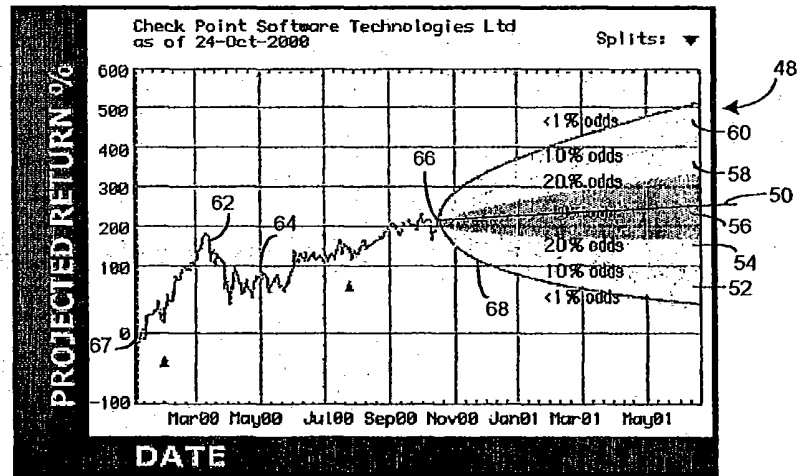


FIGURE 12

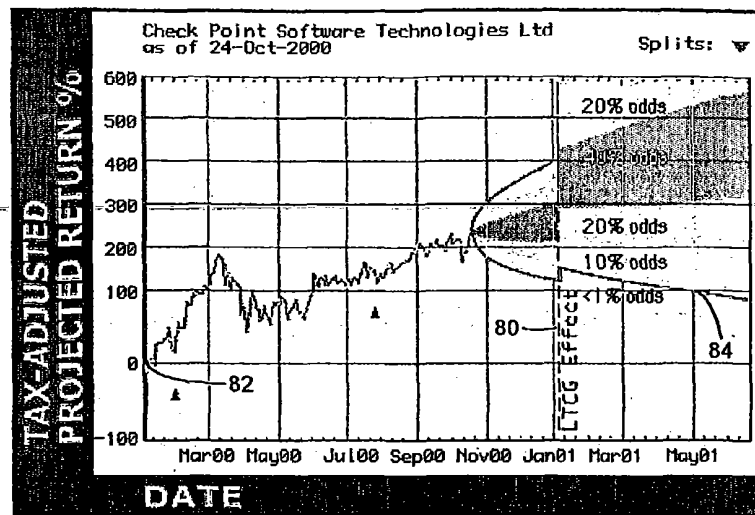


FIGURE 13

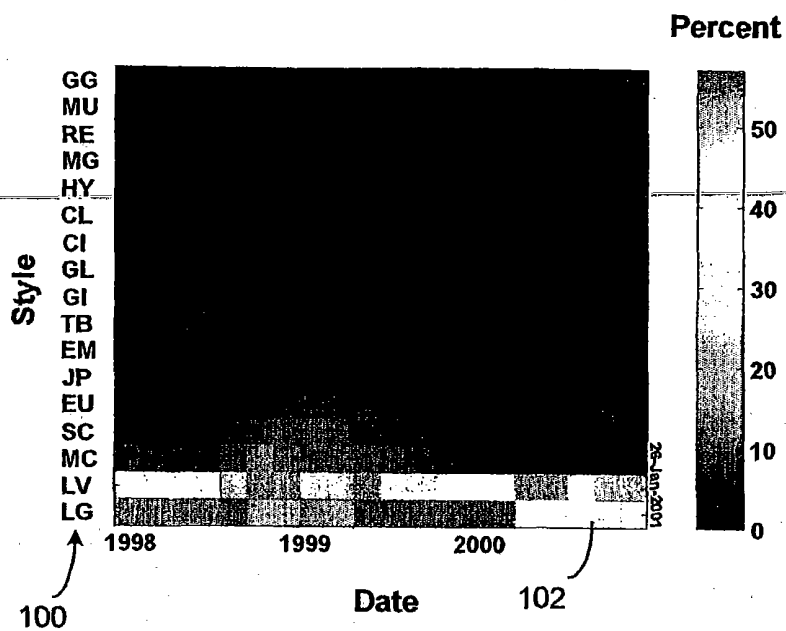


FIGURE 14

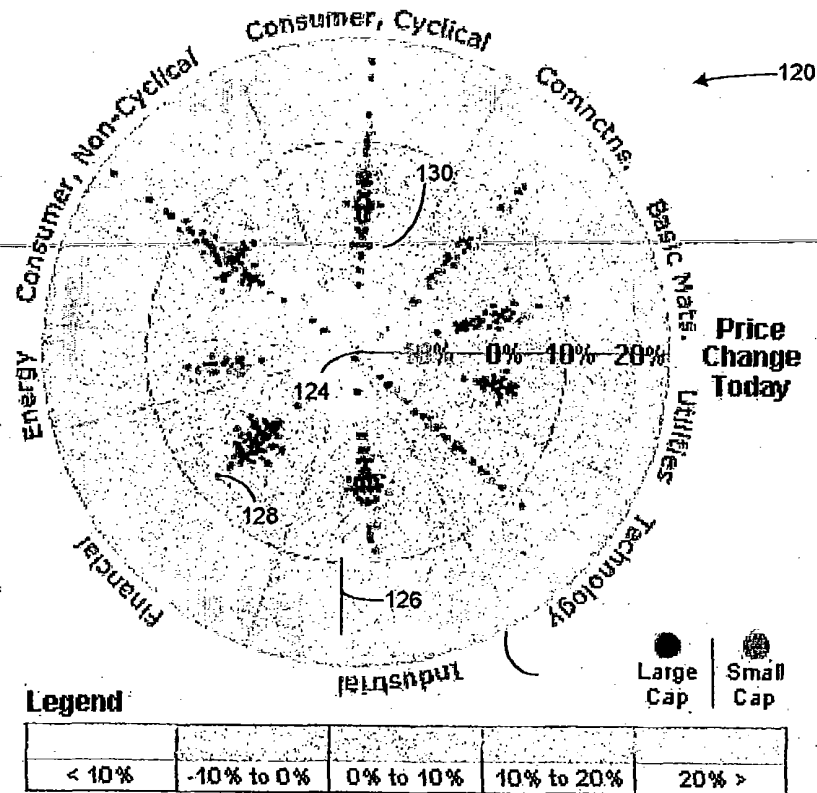


FIGURE 15

# PATENT COOPERATION TREATY

## PCT

### DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT

(PCT Article 17(2)(a), Rules 13ter.1(c) and Rule 39)


Applicant's or agent's file reference <b>11910-002W01</b>	IMPORTANT DECLARATION	Date of mailing (day/month/year) <b>29/10/2001</b>
International application No. <b>PCT/US 01/ 25753</b>	International filing date (day/month/year) <b>17/08/2001</b>	(Earliest) Priority date (day/month/year) <b>18/08/2000</b>
International Patent Classification (IPC) or both national classification and IPC <b>GO6F17/60</b>		
Applicant <b>THINKING INVESTMENTS, INC.</b>		

This International Searching Authority hereby declares, according to Article 17(2)(a), that no international search report will be established on the international application for the reasons indicated below

1. ☒ The subject matter of the international application relates to:
  - a. ☐ scientific theories.
  - b. ☐ mathematical theories
  - c. ☐ plant varieties.
  - d. ☐ animal varieties.
  - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes.
  - f. ☒ schemes, rules or methods of doing business.
  - g. ☐ schemes, rules or methods of performing purely mental acts.
  - h. ☐ schemes, rules or methods of playing games.
  - i. ☐ methods for treatment of the human body by surgery or therapy.
  - j. ☐ methods for treatment of the animal body by surgery or therapy.
  - k. ☐ diagnostic methods practised on the human or animal body.
  - l. ☐ mere presentations of information.
  - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art.
2. ☐ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:
 

☐ the description
☐ the claims
☐ the drawings
3. ☐ The failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions prevents a meaningful search from being carried out:
 

☐ the written form has not been furnished or does not comply with the standard.
 ☐ the computer readable form has not been furnished or does not comply with the standard.
4. Further comments:

Name and mailing address of the International Searching Authority  European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer <b>María Rodríguez Nóvoa</b>
--	--

## FURTHER INFORMATION CONTINUED FROM PCT/ISA 203

The claims relate to subject matter for which no search is required according to Rule 39 PCT. Given that the claims are formulated in terms of such subject matter or merely specify commonplace features relating to its technological implementation, the search examiner could not establish any technical problem which might potentially have required an inventive step to overcome. Hence it was not possible to carry out a meaningful search into the state of the art (Art. 17(2)(a)(i) and (ii) PCT; see Guidelines Part B Chapter VIII, 1-6).

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.5), should the problems which led to the Article 17(2) declaration be overcome.